

A summary of assessment tools for patients suffering from cervical spondylotic myelopathy: a systematic review on validity, reliability and responsiveness

Anoushka Singh · Lindsay Tetreault · Adrian Casey ·
Rodney Laing · Patrick Statham · Michael G. Fehlings

Received: 12 November 2012 / Revised: 18 June 2013 / Accepted: 28 July 2013 / Published online: 5 September 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract

Purpose One of the objectives of this review is to summarize the important features of a good scale. A second aim is to conduct a systematic review to identify scales that can detect the presence of cervical myelopathy and to determine their psychometric properties including validity, reliability and responsiveness.

Methods A thorough literature search was performed using MEDLINE, MEDLINE in process, EMBASE, and Cochrane Central Register of Controlled Trials. Articles were included in this study if they compared scale measurements between a control and a myelopathic patient

population or if they discussed any psychometric property of a scale.

Results An ideal scale should be one that is quantifiable, valid, sensitive, responsive and easy to perform, has high inter/intra-rater reliability, internal consistency and a suitable distribution, and is one-dimensional and relevant. In the context of cervical spondylotic myelopathy, it is essential that the scale also addresses the pathophysiology, its key signs and symptoms as well as its natural history. For the systematic review, the search yielded 5,745 citations. Of these, 37 met inclusion criteria, 10 explored the ability of a scale to detect myelopathy, 23 examined validity by assessing correlation between scales, 10 reported reliability, 8 analyzed responsiveness, and 6 discussed internal consistency. The most frequently reported scale was short form-36 ($n = 16$) followed by Nurick grade ($n = 14$), Japanese Orthopaedic Association ($n = 13$), (modified) Japanese Orthopaedic Association ($n = 7$) and grip and release test ($n = 6$). Four studies each presented results on the Cooper, Harsh and 30-m walking test.

Conclusion This review summarizes outcome measures used to assess the presence and severity of cervical myelopathy. It includes several validation studies as well as those that have reported the responsiveness and reliability of various measures.

M. G. Fehlings is Halbert Chair in Neural Repair and Regeneration, Professor of Neurosurgery at University of Toronto, ON, Canada.

A. Singh · L. Tetreault
Department of Neurosurgery, Toronto Western Hospital,
University of Toronto, 399 Bathurst St. W, 4F412,
Toronto, ON M5T 2S8, Canada

A. Casey
Department of Neurosurgery, The National Hospital for
Neurology and Neurosurgery, Queen Square,
London WC1N 3BG, UK

R. Laing
Department of Neurosurgery, Addenbrooke's Hospital,
Hills Road, Cambridge CB2 0QQ, UK

P. Statham
Department of Neurosurgery, Western General Hospital,
Crewe Road, Edinburgh EH4 2XU, Scotland, UK

M. G. Fehlings (✉)
Neural Repair and Regeneration, Neurosurgery,
University of Toronto, Krembil Neuroscience Spinal Program,
399 Bathurst St., Toronto Western Hospital,
Toronto, ON, Canada
e-mail: Michael.Fehlings@uhn.on.ca

Keywords Cervical spondylotic myelopathy ·
Outcome measures · Walking test · Impairment ·
Disability

Abbreviations

BQ	Bournemouth questionnaire
CMS	Cooper myelopathy scale
CSM	Cervical spondylotic myelopathy

CSOQ	Cervical spine outcomes questionnaire
EMS	European myelopathy score
GRT	Grip and release test
ICC	Intraclass correlation
JOACMEQ	Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire
KQ	Key question
MCS	Mental component score
MDI	Myelopathy disability index
(m)JOA	(modified) Japanese Orthopaedic Association
NDI	Neck disability index
NFDS	Neck functional disability scale
NPRS	Numeric pain rating scale
OPLL	Ossification of the posterior longitudinal ligament
PCS	Physical component score
PSFS	Patient-specific functional scale
ROC	Receiver operating curve
SF-12	Short form-12
SF-36	Short form-36
WHOQOL-Bref	World Health Organization Quality of Life-Bref

Introduction

Cervical spondylotic myelopathy (CSM) is a degenerative disease of the cervical spine and the most common cause of spinal cord dysfunction worldwide [1–3]. Both static factors, including the protrusion of osteophytic spurs, disk bulging, ossification of the posterior longitudinal ligament (OPLL) and hypertrophy of the ligamentum flavum, along with dynamic factors, such as neck extension and flexion, result in the stenosis of the spinal canal and potential compression of the neural elements [2]. Mechanical compression of the cord may lead to demyelination or focal necrosis in the posterior and lateral columns as well as loss of nerve cells in the grey matter [4]. Surgery is a common treatment option for CSM and is used to arrest its progression, prevent descent towards further neurological disability and even improve clinical status [5–8]. Many different surgical techniques have been explored, including anterior procedures such as cervical discectomy and fusion and corpectomy and posterior approaches such as laminectomy, laminectomy and fusion, and laminoplasty [9, 10].

The proper diagnosis of CSM and assessment of patient impairment and disability is essential for the implementation of appropriate treatment programs and surgical

strategies. It is sometimes the case that a patient undergoes surgery either unnecessarily or too late for optimal recovery due to non-standardized and subjective assessment [11–14]. Clinicians are increasingly using quantitative or semi-quantitative guidelines, including grading of neurological impairment, using the Ranawat scoring system, and rating of functional disability, using Japanese Orthopaedic Association (JOA) scale, Nurick grade, or the 30 m walking test [15–17]. As a result of these tools, treatment decisions are made more scientifically rather than based solely on clinical judgement. Many of these objective assessments also allow for the quantification of the severity of myelopathy. This quantification at baseline enables surgeons to predict surgical outcome, determine the necessary surgical approach for optimal results and objectively evaluate how well a patient benefited from treatment [18, 19].

The first objective of this study is to summarize important features of a good scale. The second is to conduct a systematic review of the literature to answer the following key questions (KQ): (1) what scales have the ability to detect the presence of cervical myelopathy? and (2) what are the psychometric properties of scales commonly used to assess myelopathy severity?

Qualities of an ideal scale

There are ten components that constitute an idea scale:

1. **Quantifiable:** Interval or ratio scales are more quantifiable.
2. **Suitable distribution:** the range of patients' values should be spread evenly or normally throughout the range of a scale.
3. **Validity:** how well a scale measures what it is intended to. Scales are ideally validated by correlating them with the "gold standard," something that has yet to be defined for CSM [20, 21]. A scale assessing CSM must also be validated using a CSM population [22].
4. **Sensitivity:** it is desirable for a test to record a broad range of absolute values across the population, indicating high sensitivity in detecting differences in severity between patients. A measure of sensitivity is the coefficient of variation.
5. **Responsive:** ability to detect clinically significant changes and distinguish between disease severities. Measures of responsiveness include effect sizes, area under receiver operating curves and ceiling or floor effects [23].
6. **Easy to perform:** preferably requiring no special training and having a short time frame.

7. High inter-rater and intra-rater reliability: producing the same results for repeated uses under the same conditions. Inter-observer reliability is the agreement between two or more raters, whereas intra-observer reliability is the agreement between two ratings made by a single observer on the same patient.
8. Internal consistency: a measure of the homogeneity of the scale. The Cronbach's alpha is a normalized measure of correlation between multiple components of a single scale, where a score of 1 signifies perfect correlation [24, 25].
9. One-dimensional: adds all of the components of myelopathy with no hierarchy.
10. Relevant: the measure quantifies a patient's complaints and changes as the disease progresses.

In addition, it is essential that the scale addresses the pathophysiology and the key signs and symptoms of CSM as well as its natural history. As shown in Fig. 1, the areas of increased damage in CSM are the lateral corticospinal tracts, which transmit signals for voluntary movements to the muscles, and the posterior and anterior spinocerebellar tracts, which convey sensory information on posture and movement to the cerebellum [26, 27]. These severely damaged pathways in CSM are the reasons for the presentation of its common signs and symptoms, including a broad-based spastic gait and numb, clumsy hands [28, 29]. Even though the exact anatomy and location of damage differs across the myelopathic population, this described pattern of severity can provide a rationale for selecting components of the disease to monitor. These concepts are important to remember when designing a scale intended to assess clinical and functional deficits in CSM. A measure that includes evaluation of voluntary movement, balance and coordinated activity will likely be more sensitive and relevant than one targeted towards bladder function, conscious joint position sense and temperature and pain sensation. An ideal scale should also reflect the natural history of the disease, a requirement much more difficult to meet. The onset of CSM is generally insidious and progresses in a stepwise fashion with periods of stability and periods of descent [30–33]. At the early stages of the disease, patients typically present with a mild functional gait deficit rather than an abnormal neurological examination. A scale directed towards lower limb function and sensitive enough to detect mild leg symptoms would address the natural history of this disease [34]. Along with spastic and ataxic leg complaints, the classical triad of symptoms also includes painful, stiff neck and brachialgia. These symptoms are rarely included in outcome measures directly but may indirectly be accounted for through various questions on a patient's disability.

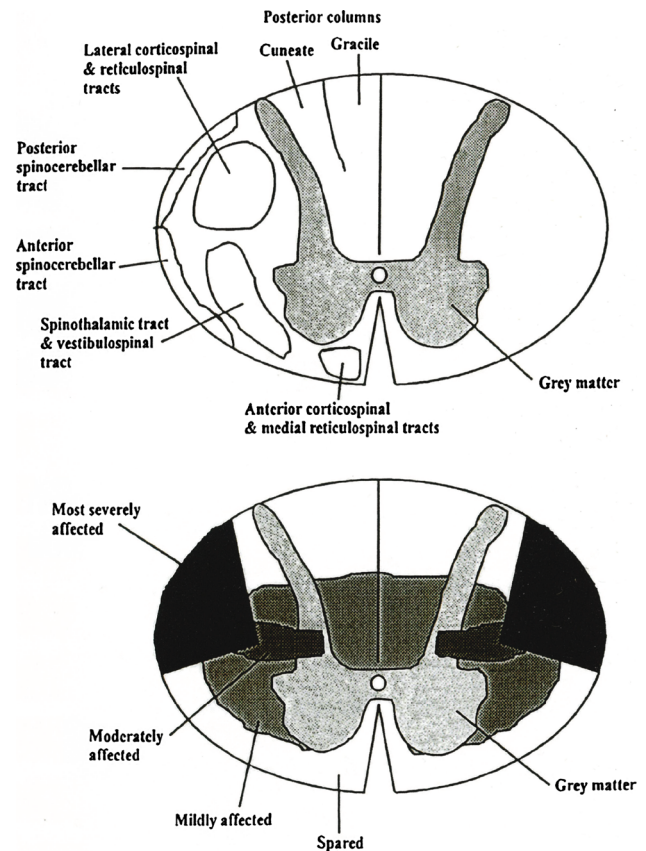


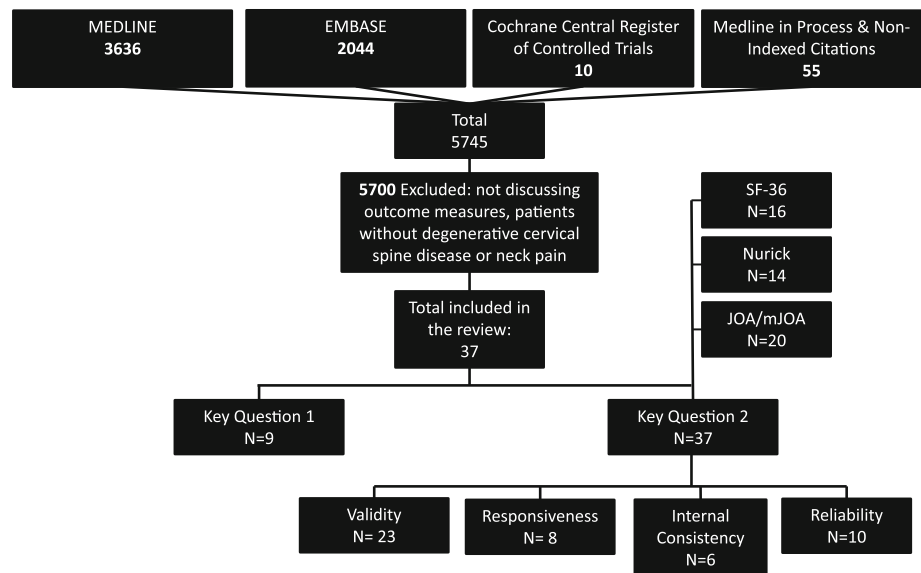
Fig. 1 Areas affected by cervical spondylotic myelopathy: essential to consider for the development of a valid, responsive and reliable outcome measure. As demonstrated by this *diagram*, the areas of increased damage are the lateral corticospinal tracts, and the posterior and anterior spinocerebellar tracts

Materials and methods

A systematic search was conducted using MEDLINE, MEDLINE in Process, EMBASE and Cochrane Central Register of Controlled Trials for literature published through November 2012. The keywords used for this search were “CSM or Ossification of the Posterior Longitudinal Ligament AND Outcome Measures AND Preoperative or Postoperative”. Two other previous searches focusing on predictors of surgical outcome and the epidemiology of the disease were also explored to identify further relevant articles. The search was limited to humans, aged 18 years or older and to English studies. Case reports, meeting abstracts/proceedings, white papers and editorials were excluded.

Studies were included for the purpose of KQ1 if they compared scale measurements between a control and a myelopathic patient population. For KQ2, studies were included if they discussed any psychometric property of a scale including its validity, reliability and responsiveness. Given that there is no published gold standard for

Fig. 2 Search strategy and detailed review process. SF-36 short form-36, JOA/mJOA (modified) Japanese orthopaedic association



the assessment of cervical myelopathy, analyses that evaluated correlation between scales were considered eligible validation studies. Studies focusing on patients with radiculopathy, rheumatoid arthritis, coexisting lumbar degeneration or neck problems were not necessarily excluded provided they discussed a scale that may be used to assess the functional status of patients with cervical myelopathy.

All abstracts and titles were reviewed independently by two of the authors and were sorted based on pre-determined inclusion and exclusion criteria. Figure 2 displays the search and review process in detail. From the included articles, the following data were extracted: author, objective, outcome measures studies, psychometric properties and correlations and conclusions. We determined which question each study answered and which psychometric properties were assessed.

Results

The total number of citations found from the three searches was 5,745. Of these, 37 met inclusion criteria: 10 answered KQ1 and 37 addressed KQ2. For KQ2, 23 studies explored validity by assessing correlation between scales, 10 examined reliability, 8 analyzed responsiveness mostly by quantifying a standardized response mean or by observing floor and ceiling effects and 6 looked at internal correlations and consistency. Scales analyzed in this review were either functional or impairment measures or were patient-reported outcomes or health assessment tools. The most frequently reported scale was SF-36 ($n = 16$) followed by Nurick grade ($n = 14$), JOA ($n = 13$), modified JOA (mJOA, $n = 7$) and the grip and release test (GRT, $n = 6$).

Four studies each presented results on the Cooper, Harsh and 30-m walking test.

Evidence that scales can detect myelopathy

Ten studies explored the ability of various scales to detect myelopathy by comparing values in a control and patient population (Table 1). Hosono et al. and Mihara et al. [35, 36] reported a significantly lower number of grip and release cycles in patients with cervical myelopathy compared to a control group (22.9 ± 8.7 vs. 32.5 ± 9.0 ; 17.3 ± 5.7 vs. 23.6 ± 4.5 cycles). Five studies explored different components of gait by performing a 30-m walking test, a 10-s step test, a foot-tapping test, a triangle step test or by simply observing a patient's stance and gait [34, 36–40]. Maezawa et al. [39] divided a 24-patient sample into three groups based on preoperative JOA score: group I consisted of patients with mild spasticity or normal function with a JOA ≥ 10 ; group II had a JOA between 7 and 9; and group III had a JOA ≤ 6 points. Compared to a control population, patients in group II or III had slower walking speed, prolonged stance phase duration, reduced stance-phase duration and a shorter stride length. In addition, all three groups had reduced knee flexion, and groups II and III had decreased plantar flexion of the ankle as compared to controls. Similarly, Kultz-Buschbeck et al. [40] also demonstrated that myelopathic patients have a slower walking velocity and a longer stance phase, as well as a wider step width and longer double support time. Singh and Crockard [34] observed similar differences in gait: mean preoperative walking time and number of steps were significantly worse for CSM patients than for controls (85.4 ± 11.2 vs. 24.3 ± 0.8 s; 74.8 ± 5.3 vs. 46.9 ± 1.2 steps).

Table 1 The ability of various measurements to detect the difference between myelopathic and control patients

Measurement	Study	Control patients	Myelopathic patients
Grip and release test (number of cycles)	Hosono et al. [35]	32.5 ± 9.0	22.9 ± 8.7
	Mihara et al. [36]	23.6 ± 4.5	17.3 ± 5.7
30-m walking test (seconds)	Maezawa et al. [39]	56.3 ± 8.3	GrI: 46.2 ± 9.4
			GrII: 37.2 ± 8.9
			GrIII: 30.2 ± 8.8
Triangle step test	Singh and Crockard [34]	24.3 ± 0.8	85.4 ± 11.2
	Mihara et al. [36]	25.4 ± 3.7	18.4 ± 5.2
Foot-tapping test	Numasawa et al. [38]	31.7 ± 6.4	23.8 ± 7.2
10 s step test	Yukawa et al. [37]	19.6 ± 3.4	10.7 ± 5.5
Gait velocity (km/h)	Kuhtz-Buschbeck et al. [40]	5.02 ± 0.81	4.28 ± 6.72
Steps	Singh and Crockard [34]	46.9 ± 1.2	74.8 ± 5.3
Step (length cm)	Maezawa et al. [39]	51.8 ± 6.2	GrI: 43.3 ± 10.5
			GrII: 35.7 ± 9.6
			GrIII: 31.3 ± 6.1
Step width (cm)	Kuhtz-Buschbeck et al. [40]	7.4 ± 1.4	9.9 ± 2.4
Stance phase duration (%)	Maezawa et al. [39]	66.7 ± 2.3	GrI: 67.7 ± 1.8
			GrII: 72.5 ± 4.0
			GrIII: 74.2 ± 5.6
Single-stance phase duration (%)	Maezawa et al. [39]	33.0 ± 2.0	GrI: 31.6 ± 3.0
			GrII: 27.2 ± 5.9
			GrIII: 22.5 ± 6.1
Stance phase	Kuhtz-Buschbeck et al. [40]	684 ± 87	756 ± 58
Double support (ms)	Kuhtz-Buschbeck et al. [40]	148 ± 28	183 ± 29
Flexion of the knee	Maezawa et al. [39]	56.6 ± 5.0	GrI: 46.5 ± 8.8
			GrII: 46.4 ± 14.2
			GrIII: 45.2 ± 7.4
Plantar flexion of the ankle (stance)	Maezawa et al. [39]	12.8 ± 5.4	GrI: 12.3 ± 4.3
			GrII: 7.5 ± 6.0
			GrIII: 8.0 ± 7.5
9-Hole peg test	Olindo et al. [41]	16.1 ± 3.1	35 ± 29.9
SF-36	King et al. [42]	PCS: 37.3 ± 10.6	PCS: 27.8 ± 8.3
		MCS: 44.4 ± 17.3	MCS: 40.5 ± 12.9

GrI patients with mild spasticity or normal, GrII JOA between 7 and 9 points, GrIII JOA ≤ 6 points, SF-36 short-form 36, PCS physical component score, MCS mental component score

The triangle step test and foot-tapping test can also detect the presence of myelopathy as demonstrated by Mihara et al. [36] and Numasawa et al. [38]. Patients with myelopathy exhibited a significantly lower number of steps on the triangular board and a lower number of foot taps in 10-s than a control (18.4 ± 5.2 vs. 25.4 ± 3.7 ; 23.8 ± 7.2 vs. 31.7 ± 6.4).

Olindo et al. [41] assessed the ability of a nine-hole peg test to quantify hand disability in a CSM population. Mean and median times to perform this test were significantly greater in the patient group than in the controls (35 ± 29.9 and 22.2 s vs. 16.1 ± 3.1 and 16.2 s).

Finally, as identified by King et al. [42], patient-reported outcome measures may differ between a myelopathic and a control population. Patients with CSM exhibited worse scores on all 8 domains of the SF-36, the physical

component score (PCS) and the mental component score (MCS) compared to age-adjusted normative data (PCS: 27.8 ± 8.3 vs. 37.3 ± 10.6 ; MCS: 40.5 ± 12.9 vs. 44.4 ± 17.3).

Psychometric properties of scales used to assess patients with cervical myelopathy

Validity

Given that there is no gold standard for the assessment of CSM, a scale's validity is typically evaluated by correlating it with other measures. Furthermore, other forms of validity including construct, convergent or discriminant, predictive and biological can be used to determine whether a scale actually measures what it is intended to. Twenty-

three studies either described previously reported correlations or one of these other forms of validation (Table 2).

Construct validity for the SF-36 was demonstrated by four studies. Brazier et al. [43] hypothesized that female patients, the elderly, those of social class IV or V and frequent health services users should have lower perceived outcome. These demographic groups did indeed display a lower SF-36 score [43]. Thakar et al. [44] proposed that mobility-related items of the SF-36 should correlate with the Nurick grade; this relationship was also true. Furthermore, King and Roberts [45], as expected, identified a significant association between Nurick grade, Cooper leg subscale and Harsh scale with physical functioning, role limitations physical, social functioning and the PCS of the SF-36. Similarly, the World Health Organization Quality of Life-Bref (WHOQOL-Bref) scale also exhibited construct validity as demonstrated by a significant relationship between preoperative physical domain scores and Nurick grade [44].

Certain subscales of the SF-36 also demonstrate discriminant and convergent validity. In a study by Baron et al. [46], role limitations physical, bodily pain, social functioning and role limitations displayed convergent and discriminant validity as their item-own correlations were greater than item-other correlations by 2 standard errors. Thakar et al. [44] confirmed convergent validity by reporting that associations between physical functioning, role limitations physical, bodily pain and PCS ranged from 0.67 to 0.74 and correlations between vitality, social functioning, role limitations emotional, mental health and MCS were from 0.67 to 0.72. Finally, predictive validity of SF-36 was proven by Latimer et al. [47] and Guilfoyle et al. [48] since the physical functioning subscale was predictive of postoperative myelopathy disability index (MDI).

The MCS and PCS of the SF-36 were highly correlated with both the preoperative and postoperative SF-12 MCS and PCS scores and the visual analog scale [49, 50]. The SF-36 physical functioning score was also highly associated with the pain and disability component of the North American Spine Society (NASS) instrument and the lower extremity domain of the Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire (JOA-CMEQ), supporting the validity of both scales [51, 52].

Multiple studies reported significant correlations with the Nurick grade: lower limb mJOA, total mJOA, JOA, Cooper myelopathy scale (CMS) of the lower and upper extremity, European myelopathy score (EMS) and poor-grade, patient-reported outcomes [20, 53–55]. Additionally, as identified by King et al. [50], there was a significant association between Nurick grade and the visual analog scale, time to trade-off and willingness to pay. Similarly, JOA was shown to be highly related to CMS of the lower

and upper extremity, EMS, time to trade-off and willingness to pay [20, 50, 55].

Casey et al. [56] identified significant correlations between MDI and spinal canal diameter, spinal cord diameter, spinal cord area, and the degree of vertical translocation of the odontoid peg through the foramen magnum, indicating biological validity. In addition, the MDI was highly associated with the EMS both preoperatively and postoperatively [20]. EMS was also significantly correlated with CMS of the upper and lower extremity [55].

To evaluate the construct validity of the neck disability index (NDI), Riddle and Stratford [57] hypothesized that patients whose work status was altered or who were undergoing litigation would have lower scores. This was proven to be true. Young et al. [58], on the other hand, speculated that “stable” patients would have NDI, patient-specific functional scale (PSFS), and numeric pain rating scale (NPRS) scores that would not change, whereas patients with a “larger” clinically, meaningfully improved would have a significant change in disability. Based on this study, NDI had poor construct validity.

Reliability

Ten studies assessed the reliability of various scales used to evaluate CSM (Table 3). Yonenobu et al. [59] conducted a study on patients with OPLL to determine the inter-rater and intra-rater reliabilities of the JOA. The interobserver reliability was high (ICC = 0.813, 95 % CI = 0.704–0.888). Table 3 displays the proportion of agreement between raters and across interviews for motor, sensory and bladder function. Intra- and inter-observer agreement was highest, 82.9 and 82.3 %, respectively, for shoulder and elbow motor function. Although Bartels et al. [60] did not validate mJOA, his study demonstrated that when researchers read a text box aloud to the patients, the kappa of the scale increased from 0.56 ± 0.11 to 0.78 ± 0.05 ($p < 0.001$).

The reliability of several gait parameters including the 10-s step test, the foot-tapping test and the 30-m walking test was described by several studies. Nakashima et al. [61] observed unsubstantial differences between and the 1st and 2nd measurement of 30 m walking test time (pre-op: 0.95, postop: 0.89), walking test steps (pre-op: -0.28 , postop: -0.55) and 10 s step test (pre-op: -0.28 , postop: -0.55). This study confirmed the findings of Singh and Crockard [34] who reported high intra-observer reliability for the 30 m walking test time and number of steps [34]. The foot-tapping test also demonstrated a high, immediate test-retest reliability in both control and myelopathy patients' groups (right side, control: $r = 0.931$, myelopathy: $r = 0.934$) [38]. Mcdermott et al. [62] performed an extensive analysis of various temporal-spatial, kinematic

Table 2 Correlations between scales and other forms of validity

Measurement	Study	Correlation with other scales	Other forms of validity
SF-36	Baron et al. [46]		Item-own scale correlations exceeded 0.40 RP, BP, SF, RE demonstrated convergent and discriminant validity (item-own > item-other correlations by 2 standard errors). PF, GH and MH had 1 item that failed to satisfy this criterion VT had 2 out of 4 items that had identical correlations with other scales. Inter-correlations between scales ranged from 0.20 to 0.62 <i>Predictive validity:</i> Preoperative SF-36 physical function was predictive of MDI postoperative ($\rho = -0.48$ to -0.60)
	Guilfoyle et al. [48]	MDI and PF (-0.76), MDI and PF (-0.56), MDI and MH (-0.39) VAS and BP (-0.57) HADS-A and PF (-0.46), HADS-A and BP (-0.58), HADS-D and MH (-0.63), HADS-D and PF (-0.63), HADS-D and BP (-0.67) Partial correlations between SF-36 PF and MH were not significant ($r = -0.01$ to 0.02)	Better functioning on the Nurick, Cooper leg subscale and Harsh scale was strongly associated with higher scores on the SF-36 PF, RP ($p < 0.001$), GH ($p < 0.005$, except for Cooper leg subscale), SF ($p < 0.003$) and PCS scales ($p < 0.012$) (<i>construct validity</i>) Moderate correlation with the mJOA score and SF-36 ($p = 0.017$ – 0.171). Strong association between lower limb JOA and SF-36 ($p < 0.006$) Low correlation between Cooper arm subscale scores and SF-36 ($p > 0.343$)
	King and Roberts [45]		
	King et al. [50]	<i>Spearman's rank correlation</i> SF-36 PCS: Standard gamble (0.119), Time trade-off (0.055), Willingness to pay ($p = 0.394$) SF-36 MCS: Standard gamble ($p = 0.052$), Time trade-off ($p = 0.181$), Willingness to pay ($p = 0.831$) Lower PCS was correlated with worse myelopathy scores on Nurick, Cooper leg subscale and Harsh scale	Validity is demonstrated by equivalent changes in the VAS, NDI and MDI Relationship between mobility-related items and Nurick (<i>construct validity</i>) Correlations between scales were fair to moderate except for PF and RE, RP and GH, RE and GH which were poor (<i>discriminant validity</i>) Associations between PF, RP, BP and PCS ranged from 0.67 – 0.74 (<i>convergent validity</i>) Associations between VT, SF, RE, MH and MCS ranged from 0.67 – 0.72 (<i>convergent validity</i>)
	King et al. [42]		
	Latimer et al. [47]		
	Thakar et al. [44]		

Table 2 continued

Measurement	Study	Correlation with other scales	Other forms of validity
WHOQOL-Bref	Thakar et al. [44]	<p>Correlation with SF-36 subscales</p> <p>PF and physical health domain (0.57)</p> <p>BP and physical health domain (0.535)</p> <p>SF and social relation domain (0.31)</p> <p>MH and psychological domain (0.35)</p> <p>PCS and physical health domain (0.511)</p> <p>MCS and psychological domain (0.23)</p> <p>MCS correlated with patient reported improvement ($p = 0.05$)</p> <p>SF-36 PCS did not correlate with patient reports of improvement</p>	Relationship between preoperative physical domain scores and Nurick (<i>construct validity</i>)
SF-12	Singh et al. [49]	<p><i>Preoperative, Postoperative</i></p> <p>SF-36 PCS (0.94, 0.97)</p> <p>SF-36 MCS (0.93, 0.96)</p>	
NASS	Stoll et al. [52]	<p><i>Preoperative, postoperative</i></p> <p>SF-36 PF (−0.75, −0.79)</p> <p>SF-36 BP (−0.65, −0.81)</p> <p>All 8 SF-36 subscales (−0.36 to −0.75, −0.39 to −0.81)</p> <p>All 8 SF-36 subscales (−0.26 to −0.40, −0.32 to −0.57)</p>	
Neurogenic symptoms			
CSOQ	BenDebba et al. [63]	Reported if >0.5	
Pain severity		ODI (−0.51), SF-36 BP (−0.50)	
Shoulder/arm pain		ODI (0.60), SF-36 BP (−0.57)	
Functional disability		ODI (0.69), SF-36 PF (−0.60), SF-36 SF (−0.64), SF-36 BP (−0.52), SF-36 PCS (−0.58)	
Psychological distress		ODI (0.50), SF-36 V (−0.59), SF-36 MH (−0.61), SF-36 SF (−0.53), SF-36 MCS (−0.65)	
Physical symptoms		ODI (0.51)	
		SF-36 PF (−0.50)	

Table 2 continued

Measurement	Study	Correlation with other scales	Other forms of validity
MDI	Casey et al. [56]	<i>Preoperative, postoperative</i>	<i>Biological validity: spinal canal diameter ($r = -0.24, p = 0.0009$), spinal cord diameter ($r = 0.29, p = 0.0009$), spinal cord area ($r = -0.34, p < 0.0001$) and the degree of vertical translocation of the odontoid peg through the foramen magnum ($r = -0.29, p = 0.0002$)</i>
	Singh and Crockard [34] Singh and Crockard [20]	HAQ (0.92, 0.53) Ranawat (0.63, 0.48) Steinbrokers (0.70, 0.50) Nurick (0.66, 0.65) <i>Preoperative, postoperative</i> EMS (0.75, 0.82), Ranawat (0.51, 0.67), Nurick (0.66, 0.71), JOA (0.56, 0.57), SF-36 (0.48, 0.35) <i>JOA with sum of left and right ($r = 0.715$)</i> R^2 (<i>preoperative, postoperative</i>) JOA (0.24, 0.17), JOACMEQ (0.24, 0.22) JOA ($r = 0.54$) <i>Preoperative Nurick (significant), IlmJOA (significant), GRT (0.55)</i> <i>Recovery rate JOA recovery rate (0.29), GRT improvement rate (0.31)</i> R^2 (<i>Preoperative, postoperative</i>) JOA (0.30, 0.17), JOACMEQ (0.40, 0.44) <i>Cutoff value of 14.5 steps</i> JOA and JOACMEQ-L were lower for patients in the low step test group <i>Improvement in JOA scores</i> Step test (HR: 1.75) <i>Improvement in JOACMEQ-L scores</i> Step test (HR: 3.31) GRT: ($r = 0.532$), Walking grade of JOA ($r = 0.841$), Total JOA ($r = 0.664$) R^2 (<i>Preoperative, postoperative</i>) JOA (0.19, 0.18), JOACMEQ (0.32, 0.31) <i>Preoperative, Postoperative</i> MDI (0.653, 0.57), Nurick (0.61, 0.69) R^2 (<i>Preoperative, postoperative</i>) JOA (0.20, 0.17), JOACMEQ (0.33, 0.26)	
Grip and release test	Hosono et al. [35] Nakashima et al. [61]		
Knee extension	Maezawa et al. [39]		
Triangle step test	Mihara et al. [36]		
10-s step test	Nakashima et al. [61]		
	Nakashima et al. [69, 69]		
30-m walking test time	Yukawa et al. [37] Nakashima et al. [61]		
30-m walking test steps	Singh and Crockard [34] Nakashima et al. [61]		

Table 2 continued

Measurement	Study	Correlation with other scales	Other forms of validity
JOACMEQ	Nikaïdo et al. [51]	<i>Preoperative, Postoperative</i>	
Upper extremity		Grip strength (0.579, 0.423)	
		GRT (0.461, 0.506)	
Cervical spine		8-Subscales of SF-36 (0.252–0.635, 0.289–0.530)	
Lower extremity		8-Subscales of SF-36 (0.307–0.532, 0.229, 0.527)	
Bladder function		8-Subscales of SF-36 (0.309–0.757, 0.436–0.882)	
Quality of life		<i>Preoperative:</i> PF (0.232), RP (0.255), GH (0.231), VT (0.257). <i>Postoperative:</i> all 8 subscales of SF-36 (0.229–0.438)	
Foot tapping test	Numasawa et al. [38]	8-Subscales of SF-36 (0.458–0.586, 0.582–0.784)	
		<i>Controls:</i> GRT (0.703)	
		<i>Patient population (pre-op):</i> JOALE ($r = 0.696$), JOA ($r = 0.662$), GRT ($r = 0.571$)	
		<i>Postoperative gain:</i> JOALE ($r = 0.431$)	
9-Hole peg test	Olindo et al. [41]	<i>Preoperative, postoperative</i>	
		mJOA (0.67, 0.52), ulmJOA (0.66, 0.57), Nurick (0.66, 0.69), Walking test (0.53, 0.45)	
Nurick	Revanappa and Rajshekhar [53]	<i>Preoperative, Postoperative:</i> lImJOA (0.901, 0.886), tmJOA (0.846, 0.862)	
		<i>Recovery rate (RR):</i> lImJOA RR (0.840), tmJOA RR (0.793)	
		<i>Change:</i> lImJOA change (0.737), tmJOA change (0.679)	
	Rajshekhar and Muliyl [54]	<i>Recovery rate:</i> PPO (whole group, 0.62; good-grade, 0.52; poor-grade 0.79)	
	Vitzthum and Dalitz [55]	JOA ($p < 0.0001$), CMS of the LE ($p < 0.0001$), EMS ($p < 0.0001$), CMS of the UE ($p < 0.05$)	
	Singh and Crockard [20]	Ranawat (0.71, 0.75), JOA (0.59, 0.51), SF-36 (0.38, 0.36)	
	King et al. [50]	<i>Spearman's rank correlation</i>	
		VAS ($p = 0.007$), Time trade-off ($p < 0.001$), Willingness to pay (0.113)	
	King et al. [70]	Not correlated with patient reports of improvement	

Table 2 continued

Measurement	Study	Correlation with other scales	Other forms of validity
Visual analog scale	Rajshekhar and Muliyl [54]	Agreement between improvement in pain scores (neck pain, arm pain, low axial pain) versus PPOs (0.01, 0.01, 0.005), NGRR (0.09, 0.10, 0.04), PCS (0.19, 0.02, 0.05), MCS (0.19, 0.02, 0.13)	
	King et al. [50]	Spearman's rank correlation SF-36 PCS ($p < 0.001$), SF-36 MCS ($p = 0.046$) Spearman's rank correlation SF-36 PCS ($p < 0.001$), SF-36 MCS ($p = 0.046$) Cuzick's nonparametric test for trend Harsh ($p = 0.001$), Cooper leg subscale ($p = 0.007$), Cooper arm subscale ($p = 0.332$), mJOA scale ($p = 0.299$) Not correlated with patient reports of improvement CMS of the LE ($p < 0.0001$), EMS ($p < 0.0001$), CMS of the UE ($p < 0.001$)	
JOA/mJOA	King et al. [70] Vitzthum and Dalitz [55]	CMS of the LE ($p < 0.0001$), EMS ($p < 0.0001$), CMS of the UE ($p < 0.001$) <i>Preoperative, postoperative</i> SF-36 (0.38, 0.37)	
	Singh and Crockard [20]	<i>Preoperative, postoperative</i> SF-36 (0.38, 0.37)	
	King et al. [50]	Cuzick's nonparametric test for trend Standard gamble ($p = 0.164$), Time trade-off ($p = 0.024$), Willingness to pay (0.010)	
EMS	King et al. [70] Vitzthum and Dalitz [55] Singh and Crockard [20]	Not correlated with patient reports of improvement CMS of the LE ($p < 0.0001$), CMS of the UE ($p < 0.001$) <i>Preoperative, postoperative</i> : Ranawat (0.61, 0.63), Nurick (0.69, 0.74), JOA (0.62, 0.72), SF-36 (0.42, 0.35)	
Prolo-score	Vitzthum and Dalitz [55]	<i>Preoperative, postoperative</i> CMS ($p < 0.05$, $p < 0.001$)	
NDI	Young et al. [58] Riddle and Stratford [57]	MCS (0.47) PCS (0.53) Change scores with GROG (0.19) Change scores with NPRS (0.61)	Poor construct validity Lower scores for NDI in patients whose work status was altered or who were undergoing litigation (<i>construct validity</i>)
PSFS	Cleland et al. [65] Young et al. [58] Cleland et al. [65]	Change scores with GROG (0.82) Change scores with NPRS (0.80)	Can distinguish stable patients from clinically meaningful improved patients

Table 2 continued

Measurement	Study	Correlation with other scales	Other forms of validity
Cooper	King et al. [50]	<i>Cuzick's nonparametric test for trend</i> Standard gamble ($p = 0.144$), Time trade-off ($p < 0.001$), Willingness to pay ($p = 0.116$)	
Leg subscale		Standard gamble ($p = 0.948$), Time trade-off ($p = 0.205$), Willingness to pay (0.116)	
Arm subscale	King et al. [70]	Not correlated with patient reports of improvement	
Harsh scale	King et al. [50]	<i>Cuzick's nonparametric test for trend</i> Standard gamble ($p = 0.024$), Time trade-off ($p = 0.087$), Willingness to pay ($p = 0.269$)	
	King et al. [70]	Not correlated with patient reports of improvement	

RP role limitations physical, *BP* bodily pain, *SF* social functioning, *RE* role limitations emotional, *PF* physical functioning; *GH* general health, *MH* mental health, *VT* vitality, *PCS* physical component score, *MCS* mental component score, *SF-36* short form-36, *NDI* neck disability index, *VAS* visual analog scale, *MDI* myelopathy disability index, *(m)JOA* (modified) Japanese Orthopaedic Association, *HADS* hospital anxiety depression scale, *SF-12* short form-12, *CSOQ* cervical spine outcomes questionnaire, *ODI* Oswestry disability index, *HAQ* health assessment questionnaire, *EMS* European myelopathy score, *JOACMEQ* Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire, *GRT* grip and release test, *l/mJOA* lower limb mJOA, *HR* hazard ratio, *ulmJOA* upper limb mJOA, *tmJOA* total mJOA, *PPO* patient-reported outcomes *CMS* Cooper myelopathy scale, *LE* lower extremity, *UE* upper extremity, *NGRR* Nurick grade recovery rate, *GROC* global rating of change, *NPRS* numeric pain rating scale

and kinetic gait parameters and quantified test–retest reliability by conducting assessments 2–7 days apart. The intraclass correlations (ICC) for all temporal-spatial parameters were high (ICC >0.9) with the exception of opposite foot contact (ICC = 0.61). Kinematic parameters displayed a wider range of reliability values ranging from 0.33 for ankle position at initial contact to 0.92 for total range pelvic obliquity. Finally, most kinetic parameters exhibited high ICCs (>0.75) except for medio-lateral group reaction force (ICC = 0.12), peak hip abductor moment (ICC = 0.54), and peak eccentric power at the knee during loading (ICC = 0.56).

As demonstrated by two studies, the GRT displays high interobserver reliability (0.989) and exhibits only small differences when measured across trials (pre-op: –0.89, postop: –0.52, right hand) [35, 61].

The reliability of six different patient-reported outcome measures was described including the NDI, PSFS, NPRS, cervical spine outcomes questionnaire (CSOQ), SF-36, and SF-12 [49, 58, 63]. In a study on patients with cervical radiculopathy, NDI and NPRS exhibited similar fair inter-observer reliability (ICC = 0.55, 0.59), where the PSFS demonstrated low reliability (ICC = 0.17). BenDebba et al. [63] evaluated this property of the CSOQ using a patient's 3-month and 12-month scores: ICCs were high for neck pain severity (0.80), shoulder arm pain severity (0.80), functional disability (0.85), psychological distress (0.82), physical symptoms other than pain (0.86), and health care utilization (0.86).

Finally, Brazier et al. [43] reported the 2-week test–retest reliability of all 8 subscales of the SF-36. The correlations were all high (>0.75) except for social functioning, role limitations physical and role limitations emotional.

Internal consistency

A scale is defined as having good internal consistency if it has a Cronbach's alpha that exceeds 0.7 or, preferably 0.8. Based on the results from several studies and as illustrated in Table 4, the SF-36 has good internal consistency with a Cronbach's alpha of 0.82 for all 36 items, >0.7 for all 8 subscales, 0.82–0.93 for the MCS and 0.89–0.92 for the PCS [43–46, 49]. Singh et al. [49] reported a lower internal consistency of 0.77 for both the MCS and PCS of the SF-12 [49]. Other reported Cronbach's alphas include MDI (0.92–0.95), EMS (0.68), JOA (0.72), and >0.85 for all domains of the WHOQOL-Bref [44] (Table 4).

Responsiveness

Eight studies reported on a scale's responsiveness (Table 5). The effect sizes for the SF-36, WHOQOL-Bref,

Table 3 Interobserver and test–retest reliability

Measurement	Study	Interobserver reliability	Test–retest reliability
JOA (total score)	Yonenobu et al. [59]	ICC = 0.813	ICC = 0.826
Motor function		K (proportion of agreement)	K (proportion of agreement)
Fingers		$\kappa = 0.534$ (77.7 %)	$\kappa = 0.678$ (73.1 %)
Shoulder and elbow		$\kappa = 0.305$ (82.3 %)	$\kappa = 0.501$ (82.9 %)
Lower extremity		$\kappa = 0.488$ (62.3 %)	$\kappa = 0.547$ (62.2 %)
Sensory function			
Upper extremity		$\kappa = 0.421$ (67.7 %)	$\kappa = 0.510$ (64.7 %)
Trunk		$\kappa = 0.579$ (78.5 %)	$\kappa = 0.537$ (75.7 %)
Lower extremity		$\kappa = 0.339$ (62.3 %)	$\kappa = 0.436$ (57.1 %)
Bladder function		$\kappa = 0.469$ (75.4 %)	$\kappa = 0.643$ (75.1 %)
Translated version of mJOA	Bartels et al. [60]	$\kappa = 0.56 \pm 0.11$ $\kappa = 0.78 \pm 0.05$ (after instructions)	
Grip and release test	Hosono et al. [35] Nakashima et al. [61]	ICC = 0.989 (mean) Not reported	Not reported differences (SD) between 1st and 2nd measurements <i>Right hand</i> <i>Preoperative</i> : –0.89 (1.91) <i>Postoperative</i> : –0.52 (1.74) <i>Left hand</i> <i>Preoperative</i> : –0.46 (1.68) <i>Postoperative</i> : 0.02 (2.00)
10 s step test	Nakashima et al. [61]	Not reported	Differences (SD) between 1st and 2nd measurements <i>Preoperative</i> : –0.28 (1.04) <i>Postoperative</i> : –0.55 (0.87)
Foot tapping test	Numasawa et al. [38]	Not reported	<i>Immediate test–retest</i> <i>Right side (control)</i> $R = 0.931, p < 0.0001$ <i>Left side (control)</i> $R = 0.899, p < 0.0001$ <i>Right side (myelopathy)</i> $R = 0.934, p < 0.0001$ <i>Left side (myelopathy)</i> $R = 0.899, p < 0.0001$
30 m walking test	Nakashima et al. [61] Singh and Crockard [34]	Not reported	Differences (SD) between 1st and 2nd measurements <i>Preoperative</i> : 0.95 (1.66) <i>Postoperative</i> : 0.89 (1.33) $p = 0.995$ (low variability between trials)
Gait parameters ^a	McDermott et al. [62]	Not reported	Assessments were 2–7 days apart

Table 3 continued

Measurement	Study	Interobserver reliability	Test–retest reliability
Temporal spatial			
Cadence			ICC = 0.99
Double support			ICC = 0.99
Foot off			ICC = 0.97
Single support			ICC = 0.96
Step length			ICC = 0.99
Stride length			ICC = 0.99
Stride time			ICC = 0.99
Gait speed			ICC = 0.99
Kinematic			
Total hip range in sagittal plane			ICC = 0.99
Kinetic			
Second peak, antero-posterior GRF			ICC = 0.95
Hip concentric power, terminal stance			ICC = 0.95
NDI	Young et al. [58]	Not reported	Test–retest over 4 weeks ICC = 0.55
PSFS	Young et al. [58]	Not reported	Test–retest over 4 weeks ICC = 0.17
NPRS	Young et al. [58]	Not reported	Test–retest over 4 weeks ICC = 0.59
CSOQ	BenDebba et al. [63]	Not reported	Test–retest (3–6 month scores)
Neck pain			
Shoulder-arm pain			ICC = 0.80
Functional disability			ICC = 0.85
Psychological distress			ICC = 0.82
Physical symptoms			ICC = 0.86
Health care utilization			ICC = 0.75
SF-36	Brazier et al. [43]	Not reported	Test–retest (2 week interval)
Physical functioning			$R = 0.81$
Social functioning			$R = 0.60$
Role limitations (P)			$R = 0.69$
Role limitations (E)			$R = 0.63$
Bodily pain			$R = 0.78$
Mental health			$R = 0.75$
Vitality			$R = 0.80$
General health perception			$R = 0.80$

JOA Japanese Orthopaedic Association, ICC intraclass correlation, K kappa coefficient of Kraemer, SD standard deviation, NDI neck disability index, PSFS patient-specific functional scale, NPRS numeric pain rating scale, P physical, E emotional

^a Only the gait parameters with ICC >0.95 are reported

Bournemouth questionnaire (BQ), NDI, neck functional disability scale (NFDS), NPRS and MDI have been defined in the literature [44, 46, 48, 56, 58, 64, 65]. Four studies referenced Cohen's classification of effect size that defines 0.2 as a small group change, 0.5 as moderate and 0.8 as large. According to these criterion, three composite measures of the CSOQ are considered highly responsive (neck pain = 0.71–1.15, functional disability 0.9–1.0,

physical symptoms 0.63–0.95), two moderately responsive (shoulder-arm pain = 0.68–0.79, psychological distress = 0.58–0.72) and one poorly responsive (health care utilization = 0.21–0.36) [63]. For the SF-36, Baron et al. [46] identified a significant change in the value of 5 scores postoperatively: effect sizes were –0.70 for energy/vitality, –0.55 for mental health, –0.52 for social functioning, –0.62 for bodily pain and –0.43 for physical functioning.

Table 4 Internal consistency determined by Cronbach's alpha

Measurement	Study	Cronbach's Alpha
SF-36	Baron et al. [46], King and Roberts [45], Thakar et al. [44]	PF (0.94, 0.91, 0.86) RP (0.84, 0.81, 0.86) BP (0.89, 0.89, 0.85) GH (0.79, 0.79, 0.86) RE (0.93, 0.83, 0.87) SF (0.84, 0.84, 0.85) MH (0.78, 0.89, 0.85) VT (0.81, 0.83, 0.84)
	King and Roberts [45], Singh et al. [49], Thakar et al. [44], Singh and Crockard [20]	MCS (0.92, 0.93, 0.82) PCS (0.92, 0.89, 0.89) <i>Preoperative, postoperative</i> All 36-items (0.82, 0.86)
SF-12	Singh et al. [49]	PCS (0.77), MCS (0.77)
WHOQOL-Bref	Thakar et al. [44]	PH (0.86), PS (0.86), SR (0.87), EN (0.87)
MDI	Casey et al. [56] Singh and Crockard [20]	0.95 <i>Preoperative, postoperative</i> 11 Scores (0.92, 0.96) 4 Categories (0.76, 0.81)
EMS	Singh and Crockard [20]	<i>Preoperative</i> : 0.68 <i>Postoperative</i> : 0.77
JOA	Singh and Crockard [20]	<i>Preoperative, postoperative</i> 6 Scores: (0.72, 0.73) 4 Categories (0.66, 0.65)

SF-36 short form-36, PF physical functioning, RP role limitations physicals, BP bodily pain, GH general health, RE role limitations emotional, SF social functioning, MH mental health, VT vitality, MCS mental component score, PCS physical component score, SF-12 short form-12, WHOQOL-Bref World Health Organization Quality of Life-Bref, PH physical quality of life, PS psychological quality of life, SR social relation, EN environmental quality of life, MDI myelopathy disability index, EMS European myelopathy score, JOA Japanese Orthopaedic Association

Thakar et al. [44] and Guilfoyle et al. [48] reported similar effect sizes for mental health (0.54 for both studies) and higher effect sizes for physical functioning (0.78; 0.86). The responsiveness of emotional role was moderate (ES = 0.62) [44]. For bodily pain, Thakar et al. [44] and Guilfoyle et al. [48] observed effect sizes of 0.80 and 0.65, respectively. The domains for WHOQOL-Bref, on the other hand, were generally less responsive than the SF-36 subscales, with the exception of the physical domain (ES = 0.68) [44]. Effect sizes were 0.03, 0.39 and 0.45 for social, psychological and environment, respectively. As described by Bolton [64], all BQ subscales (0.82–1.94), BQ total score (1.67) and NRS pain (1.02–1.60) were all highly responsive to change and NDI and NFDS were moderately responsive. The effect size for MDI was reported by Casey et al. [56] as 0.39.

Other studies described responsiveness by calculating the area under a receiver operator curve (ROC) [58, 65]. This curve plots the sensitivity against 1-specificity to describe a scale's ability to differentiate between two groups of patients. Young et al. [58] divided their patient sample into

four groups based on whether they were unchanged, stable, "smaller" clinically improved or "larger" clinically improved following surgery. The area under the ROC for NDI, PSFS and NPRS were 0.74, 0.71 and 0.72, respectively, when differentiating between stable and "larger" clinically improved. Cleland et al. [65], on the other hand, reported a much lower area of 0.57 for NDI (stable and improved patients) and a higher area of 0.99 for the PSFS.

Three studies identified significant floor and ceiling effects for subscales of the SF-36. Baron et al. [46], Guilfoyle et al. [48], and Thakar et al. [44] reported a high floor effect for physical role (62.8, 50.9 and 64.3 %) and emotional role (35.1, 20, 71.4 %). Studies have reported floor effects of 15.4 % for physical functioning, 30 % for bodily pain and 17.1 % for social functioning and have identified ceiling effects of 22.9 % for social functioning, 20.4 % for physical role and 51.8 % for emotional role [44, 46, 48]. Casey et al. [56] noted a floor effect for the MDI as 27 out of their 194 sample of surgical patients had maximum disability. Although not explicitly examined in the literature, previous studies have speculated that the mJOA

Table 5 Responsiveness given by effect size, area under a receiver operating curve or floor and ceiling effects

Measurement	Study	Effect size or AUC	Floor, ceiling effect
SF-36	Baron et al. [46]	PF (−0.43), BP (−0.62), SF (−0.52)	PF (15.4 %, NS),
		MH (−0.55), VT (−0.70)	RP (62.8 %, NS)
			RE (35.1 %, 51.8 %)
	Guilfoyle et al. [48]	<i>Standardised response means</i>	RP (50.9 %, 20.4 %)
		PF (0.86), BP (0.65), MH (0.54)	SF (NS, 22.9 %)
			RE (29.0, 52.1 %)
	Thakar et al. [44]	PF (0.78), RP (0.49), BP (0.80),	RP (64.3 %, NS)
		GH (0.39), VT (0.53), SF (0.65),	BP (30 %, NS)
		RE (0.62), MH (0.54)	SF (17.1 %, NS)
		RE (71.4 %, NS)	
CSOQ	BenDebba et al. [63]	<i>3-month/6-month</i>	
Neck pain severity		(0.71, 1.15)	
Shoulder/arm pain severity		(0.68, 0.79)	
Functional disability		(0.90, 1.00)	
Psychological distress		(0.72, 0.58)	
Physical symptoms		(0.63, 0.95)	
Healthcare utilization		(0.21, 0.36)	
BQ total score	Bolton [64]	1.67	
Pain intensity		1.94	
Function-ADL		1.47	
Function-social		1.14	
Anxiety		1.25	
Depression		0.82	
Fear-avoidance		0.83	
Self-locus of control		1.67	
NDI	Bolton [64]	0.80	
	Young et al. [58]	<i>AUC: improved and stable patients</i>	
		(0.57)	
		<i>AUC: stable and “smaller” clinically improved</i> (0.61)	
		<i>AUC: stable and “larger” clinically improved</i> (0.74)s	
PSFS	Cleland et al. [65]	<i>AUC: improved and stable patients</i>	
		(0.99)	
	Young et al. [58]	<i>AUC: stable and “smaller” clinically improved</i> (0.66)	
		<i>AUC: stable and “larger” clinically improved</i> (0.71)	
NPRS	Cleland et al. [65]	1.02	
		1.60	
		1.23	
	Young et al. [58]	<i>AUC: stable and “smaller” clinically improved</i> (0.61)	
		<i>AUC: stable and “larger” clinically improved</i> (0.72)	

Table 5 continued

Measurement	Study	Effect size or AUC	Floor, ceiling effect
NFDS	Bolton [64]	0.58	
WHOQOL-Bref	Thakar et al. [44]		
Physical		0.68	
Psychological		0.39	
Social		0.03	
Environment		0.45	
MDI	Casey et al. [56]	0.39	Floor effect (27/194) surgical patients have maximum disability

Cohen's effect size

Area under a receiver operating curve (AUC) reflecting the ability of a scale to distinguish between two groups

SF-36 short form-36, *PF* physical functioning, *BP* bodily pain, *SF* social functioning, *MH* mental health, *VT* vitality, *RP* role limitations physical, *RE* role limitations emotional, *GH* general health, *CSOQ* cervical spine outcomes questionnaires, *BQ* Bournemouth questionnaire, *NDI* neck disability index, *PSFS* patient-specific functional scale, *NFDS* neck functional disability scale, *WHOQOL-Bref* World Health Organization Quality of Life-Bref, *MDI* myelopathy disability index

has a ceiling effect. In a study by Fehlings et al. [66], the degree of improvement on the mJOA scale was directly associated with baseline severity score, with severe patients showing the greatest change in mJOA. As a mild patient has less room for improvement as their baseline score is closer to the perfect score of 18, they are expected to show smaller changes in mJOA score.

Discussion

It is evident from this review that there are numerous measurement tools that may be used to assess patients with cervical myelopathy. Singh et al. [67] reported the results of a survey in which clinicians identified that, although it is essential to quantify functional disability in CSM patients, the specific measurements that may do this are underused or not ideal [67]. Even though some scales are more commonly used in global practice, there still remains a lack of a gold standard outcome measure that can assess disease severity over time and objectively determine a patient's improvement following intervention. This prevents the establishment of standard quantitative guidelines used to direct and implement appropriate treatment programs and makes it more challenging to accurately predict surgical outcome. It is therefore essential to develop or define a gold standard to determine the ideal timing of surgical intervention, thereby optimizing the clinical management of CSM populations.

Tetreault et al. [68], in a systematic review on important predictors of outcome, identified that results may differ depending on what scale was used to evaluate outcome. For example, when outcome was assessed using the Nurick grade, its association with various predictors was less conclusive: the direction of the relationship between

preoperative condition and Nurick grade was unclear as was the hypothesis that duration of symptoms was correlated with a worse outcome. The authors speculated these findings were likely due to limitations in the Nurick grade as it is insensitive and largely weighted towards employment and lower limb function. The results from this study demonstrate that even the most commonly used outcome measures have significant flaws that may prevent them from correctly identifying certain relationships.

This systematic review provides an extensive summary of the outcome measures used for the functional assessment and quality of life evaluation of CSM. Some of these scales, such as the mJOA, Nurick and 30-m walking test have gained acceptance and are widely used whereas others have not yet broken into clinical practice. This study has explored the psychometric properties and the utility of existing scales in an attempt to propose an ideal measure.

Unfortunately, we were not able to conclusively identify the ideal scale for evaluating patients with cervical myelopathy as we believe it does not exist. An ideal outcome measure would be one that incorporates all ten described qualities and also plays to the pathophysiology, signs and symptoms and natural history of the disease. Given that CSM typically affects the lateral corticospinal and the spinocerebellar tracts, a scale that assesses voluntary movement, balance and coordinated activity will likely be more sensitive and relevant than one that does not. In addition, a scale that can detect mild symptoms will address the insidious and slowly progressive nature of the disease. It clear from Tables 1 through 5 that certain scales can detect the presence of myelopathy, are valid, highly reliable, responsive to change and have a high internal consistency. Despite this, there are definite limitations to all measures reported. Based on these findings, we propose that functional measures such as the MDI, mJOA and Nurick grade

should be used in combination with more sensitive and quantitative measures including various walking tests or the grip and release test. In addition, the use a quality life measures such as the SF-36 may help determine the impact of myelopathy on a patient's mental and physical status. The use of a wider range of functional and impairment tests may help better define optimal treatment interventions as well as the prognostic value of certain clinical and imaging factors. In addition, we hope this review will direct future research in the form of a high quality clinimetric study exploring the quality of performance of key measures.

Conflict of interest None.

References

- Tracy JA, Bartleson JD (2010) Cervical spondylotic myelopathy. *Neurologist* 16:176–187
- Baptiste DC, Fehlings MG (2006) Pathophysiology of cervical myelopathy. *Spine J* 6:190S–197S. doi:10.1016/j.spinee.2006.04.024
- Nurick S (1972) The pathogenesis of the spinal cord disorder associated with cervical spondylosis. *Brain* 95:87–100
- Matz PG, Anderson PA, Holly LT, Groff MW, Heary RF, Kaiser MG, Mummaneni PV, Ryken TC, Choudhri TF, Vresilovic EJ, Resnick DK (2009) The natural history of cervical spondylotic myelopathy. *J Neurosurg Spine* 11:104–111. doi:10.3171/2009.1.spine08716
- Fehlings MG, Arvin B (2009) Surgical management of cervical degenerative disease: the evidence related to indications, impact, and outcome. *J Neurosurg Spine* 11:97–100. doi:10.3171/2009.5.spine09210
- Kaminsky SB, Clark CR, Traynelis VC (2004) Operative treatment of cervical spondylotic myelopathy and radiculopathy. A comparison of laminectomy and laminoplasty at five year average follow-up. *Iowa Orthop J* 24:95–105
- Kiris T, Kilincer C (2008) Cervical spondylotic myelopathy treated by oblique corpectomy: a prospective study. *Neurosurgery* 62:674–681. doi:10.1227/01.neu.0000317316.56235.a7
- Papadopoulos CA, Katonis P, Papagelopoulos PJ, Karampekios S, Hadjipavlou AG (2004) Surgical decompression for cervical spondylotic myelopathy: correlation between operative outcomes and MRI of the spinal cord. *Orthopedics* 27:1087–1091
- Mattei TA, Goulart CR, Milano JB, Dutra LP, Fasset DR (2011) Cervical spondylotic myelopathy: pathophysiology, diagnosis, and surgical techniques. *ISRN Neurol* 2011:463729. doi:10.5402/2011/463729
- Mummaneni PV, Kaiser MG, Matz PG, Anderson PA, Groff MW, Heary RF, Holly LT, Ryken TC, Choudhri TF, Vresilovic EJ, Resnick DK (2009) Cervical surgical techniques for the treatment of cervical spondylotic myelopathy. *J Neurosurg Spine* 11:130–141. doi:10.3171/2009.3.spine08728
- Lunsford LD, Bissonette DJ, Jannetta PJ, Sheptak PE, Zorub DS (1980) Anterior surgery for cervical disc disease. Part 1: treatment of lateral cervical disc herniation in 253 cases. *J Neurosurg* 53:1–11. doi:10.3171/jns.1980.53.1.0001
- Harsh GR IV, Sybert GW, Weinstein PR, Ross DA, Wilson CB (1987) Cervical spine stenosis secondary to ossification of the posterior longitudinal ligament. *J Neurosurg* 67:349–357. doi:10.3171/jns.1987.67.3.0349
- Allen KL (1952) Neuropathies caused by bony spurs in the cervical spine with special reference to surgical treatment. *J Neurol Neurosurg Psychiatry* 15:20–36
- Bakay L, Cares HL, Smith RJ (1970) Ossification in the region of the posterior longitudinal ligament as a cause of cervical myelopathy. *J Neurol Neurosurg Psychiatry* 33:263–268
- Ranawat CS, O'Leary P, Pellicci P, Tsairis P, Marchisello P, Dorr L (1979) Cervical spine fusion in rheumatoid arthritis. *J Bone Joint Surg Am* 61:1003–1010
- Steinbrocker O, Traeger CH, Batterman RC (1949) Therapeutic criteria in rheumatoid arthritis. *J Am Med Assoc* 140:659–662
- Hirabayashi K, Watanabe K, Wakano K, Suzuki N, Satomi K, Ishii Y (1983) Expansive open-door laminoplasty for cervical spinal stenotic myelopathy. *Spine (Phila Pa 1976)* 8:693–699
- Laing RJ (2000) Measuring outcome in neurosurgery. *Br J Neurosurg* 14:181–184
- Holly LT, Matz PG, Anderson PA, Groff MW, Heary RF, Kaiser MG, Mummaneni PV, Ryken TC, Choudhri TF, Vresilovic EJ, Resnick DK (2009) Clinical prognostic indicators of surgical outcome in cervical spondylotic myelopathy. *J Neurosurg Spine* 11:112–118. doi:10.3171/2009.1.spine08718
- Singh A, Crockard HA (2001) Comparison of seven different scales used to quantify severity of cervical spondylotic myelopathy and post-operative improvement. *J Outcome Meas* 5:798–818
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
- Wade DT (1992) Measurement in neurological rehabilitation. *Curr Opin Neurol Neurosurg* 5:682–686
- Deyo RA, Diehr P, Patrick DL (1991) Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials* 12:142S–158S
- Cronbach LJ, Meehl PE (1955) Construct validity in psychological tests. *Psychol Bull* 52:281–302
- Kraemer HC (1980) Extension of the kappa coefficient. *Biometrics* 36:207–216
- Ogino H, Tada K, Okada K, Yonenobu K, Yamamoto T, Ono K, Namiki H (1983) Canal diameter, anteroposterior compression ratio, and spondylotic myelopathy of the cervical spine. *Spine (Phila Pa 1976)* 8:1–15
- Levine DN (1997) Pathogenesis of cervical spondylotic myelopathy. *J Neurol Neurosurg Psychiatry* 62:334–340
- Arnold JG Jr (1955) The clinical manifestations of spondylolordosis (spondylosis) of the cervical spine. *Ann Surg* 141:872–889
- Brain WR, Northfield D, Wilkinson M (1952) The neurological manifestations of cervical spondylosis. *Brain* 75:187–225
- Clark CR (1988) Cervical spondylotic myelopathy: history and physical findings. *Spine* 13:847–849
- Hunt WE (1980) Cervical spondylosis: natural history and rare indications for surgical decompression. *Clin Neurosurg* 27:466–480
- Lees F, Turner JW (1963) Natural history and prognosis of cervical spondylosis. *Br Med J* 2:1607–1610
- Clarke E, Robinson PK (1956) Cervical myelopathy: a complication of cervical spondylosis. *Brain* 79:483–510
- Singh A, Crockard HA (1999) Quantitative assessment of cervical spondylotic myelopathy by a simple walking test. *Lancet* 354:370–373. doi:10.1016/s0140-6736(98)10199-x
- Hosono N, Sakaura H, Mukai Y, Kaito T, Makino T, Yoshikawa H (2008) A simple performance test for quantifying the severity of cervical myelopathy [Erratum appears in *J Bone Joint Surg Br*. 2008 Nov; 90(11):1534]. *J Bone Joint Surg Br* 90:1210–1213. doi:10.1302/0301-620X.90B9.20459

36. Mihara H, Kondo S, Murata A, Ishida K, Niimura T, Hachiya M (2010) A new performance test for cervical myelopathy: the triangle step test. *Spine (Phila Pa 1976)* 35:32–35. doi:[10.1097/BRS.0b013e3181b839b0](https://doi.org/10.1097/BRS.0b013e3181b839b0)
37. Yukawa Y, Kato F, Ito K, Horie Y, Nakashima H, Masaaki M, Ito ZY, Wakao N (2009) “Ten second step test” as a new quantifiable parameter of cervical myelopathy. *Spine (Phila Pa 1976)* 34:82–86. doi:[10.1097/BRS.0b013e31818e2b19](https://doi.org/10.1097/BRS.0b013e31818e2b19)
38. Numasawa T, Ono A, Wada K, Yamasaki Y, Yokoyama T, Aburakawa S, Takeuchi K, Kumagai G, Kudo H, Umeda T, Nakaji S, Toh S (2012) Simple foot tapping test as a quantitative objective assessment of cervical myelopathy. *Spine* 37:108–113. doi:[10.1097/BRS.0b013e31821041f8](https://doi.org/10.1097/BRS.0b013e31821041f8)
39. Maezawa Y, Uchida K, Baba H (2001) Gait analysis of spastic walking in patients with cervical compressive myelopathy. *J Orthop Sci* 6:378–384
40. Kutz-Buschbeck JP, Johnk K, Mader S, Stolze H, Mehdorn M (1999) Analysis of gait in cervical myelopathy. *Gait Posture* 9:184–189
41. Olindo S, Signate A, Richech A, Cabre P, Catonne Y, Smadja D, Pascal-Mousselard H (2008) Quantitative assessment of hand disability by the Nine-Hole-Peg test (9-HPT) in cervical spondylotic myelopathy. *J Neurol Neurosurg Psychiatry* 79:965–967. doi:[10.1136/jnnp.2007.140285](https://doi.org/10.1136/jnnp.2007.140285)
42. King Jr JT, McGinnis KA, Roberts MS (2003) Quality of life assessment with the medical outcomes study short form-36 among patients with cervical spondylotic myelopathy. *Neurosurgery* 52:113–120 (discussion 121)
43. Brazier JE, Harper R, Jones NM, O’athain A, Thomas KJ, Usherwood T, Westlake L (1992) Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 305:160–164
44. Thakar S, Christopher S, Rajshekhar V (2009) Quality of life assessment after central corpectomy for cervical spondylotic myelopathy: comparative evaluation of the 36-Item Short Form Health Survey and the World Health Organization Quality of Life-Bref. *Clinical article. J Neurosurg Spine* 11:402–412. doi:[10.3171/2009.4.SPINE08749](https://doi.org/10.3171/2009.4.SPINE08749)
45. King JT Jr, Roberts MS (2002) Validity and reliability of the Short Form-36 in cervical spondylotic myelopathy. *J Neurosurg* 97:180–185
46. Baron R, Elashaal A, Germon T, Hobart J (2006) Measuring outcomes in cervical spine surgery: think twice before using the SF-36. *Spine* 31:2575–2584. doi:[10.1097/01.brs.0000240694.83621.52](https://doi.org/10.1097/01.brs.0000240694.83621.52)
47. Latimer M, Haden N, Seeley HM, Laing RJ (2002) Measurement of outcome in patients with cervical spondylotic myelopathy treated surgically. *Br J Neurosurg* 16:545–549
48. Guilfoyle MR, Seeley H, Laing RJ (2009) The Short Form 36 health survey in spine disease—validation against condition-specific measures. *Br J Neurosurg* 23:401–405. doi:[10.1080/02688690902730731](https://doi.org/10.1080/02688690902730731)
49. Singh A, Gnanalingham K, Casey A, Crockard A (2006) Quality of life assessment using the Short Form-12 (SF-12) questionnaire in patients with cervical spondylotic myelopathy: comparison with SF-36. *Spine* 31:639–643. doi:[10.1097/01.brs.0000202744.48633.44](https://doi.org/10.1097/01.brs.0000202744.48633.44)
50. King JT Jr, Tsevat J, Moosy JJ, Roberts MS (2004) Preference-based quality of life measurement in patients with cervical spondylotic myelopathy. *Spine* 29:1271–1280
51. Nikaido T, Kikuchi S-I, Yabuki S, Otani K, Konno S-I (2009) Surgical treatment assessment using the Japanese orthopedic association cervical myelopathy evaluation questionnaire in patients with cervical myelopathy: a new outcome measure for cervical myelopathy. *Spine* 34:2568–2572
52. Stoll T, Huber E, Bachmann S, Baumeler HR, Mariacher S, Rutz M, Schneider W, Spring H, Aeschlimann A, Stucki G, Steiner W (2004) Validity and sensitivity to change of the NASS questionnaire for patients with cervical spine disorders. *Spine (Phila Pa 1976)* 29:2851–2855
53. Revanappa KK, Rajshekhar V (2011) Comparison of Nurick grading system and modified Japanese Orthopaedic Association scoring system in evaluation of patients with cervical spondylotic myelopathy. *Eur Spine J* 20:1545–1551. doi:[10.1007/s00586-011-1773-y](https://doi.org/10.1007/s00586-011-1773-y)
54. Rajshekhar V, Muliyl J (2007) Patient perceived outcome after central corpectomy for cervical spondylotic myelopathy. *Surgical Neurology* 68:185–190. doi:[10.1016/j.surneu.2006.10.071](https://doi.org/10.1016/j.surneu.2006.10.071)
55. Vitzthum H-E, Dalitz K (2007) Analysis of five specific scores for cervical spondylogenic myelopathy. *Eur Spine J* 16:2096–2103
56. Casey AT, Bland JM, Crockard HA (1996) Development of a functional scoring system for rheumatoid arthritis patients with cervical myelopathy. *Ann Rheum Dis* 55:901–906
57. Riddle DL, Stratford PW (1998) Use of generic versus region-specific functional status measures on patients with cervical spine disorders. *Phys Ther* 78:951–963
58. Young IA, Cleland JA, Michener LA, Brown C (2010) Reliability, construct validity, and responsiveness of the neck disability index, patient-specific functional scale, and numeric pain rating scale in patients with cervical radiculopathy. *Am J Phys Med Rehabil* 89:831–839. doi:[10.1097/PHM.0b013e3181ec98e6](https://doi.org/10.1097/PHM.0b013e3181ec98e6)
59. Yonenobu K, Abumi K, Nagata K, Taketomi E, Ueyama K (2001) Interobserver and intraobserver reliability of the Japanese orthopaedic association scoring system for evaluation of cervical compression myelopathy. *Spine (Phila Pa 1976)* 26:1890–1894 (discussion 1895)
60. Bartels RH, Verbeek AL, Benzel EC, Fehlings MG, Guiot BH (2010) Validation of a translated version of the modified Japanese orthopaedic association score to assess outcomes in cervical spondylotic myelopathy: an approach to globalize outcomes assessment tools. *Neurosurgery* 66:1013–1016. doi:[10.1227/01.ane.0000368391.79314.6f](https://doi.org/10.1227/01.ane.0000368391.79314.6f)
61. Nakashima H, Yukawa Y, Ito K, Machino M, Kanbara S, Morita D, Imagama S, Hamajima N, Ishiguro N, Kato F (2011) Validity of the 10-s step test: prospective study comparing it with the 10-s grip and release test and the 30-m walking test. *Eur Spine J* 20:1318–1322. doi:[10.1007/s00586-011-1733-6](https://doi.org/10.1007/s00586-011-1733-6)
62. McDermott A, Bolger C, Keating L, McEvoy L, Meldrum D (2010) Reliability of three-dimensional gait analysis in cervical spondylotic myelopathy. *Gait Posture* 32:552–558. doi:[10.1016/j.gaitpost.2010.07.019](https://doi.org/10.1016/j.gaitpost.2010.07.019)
63. BenDebba M, Heller J, Ducker TB, Eisinger JM (2002) Cervical spine outcomes questionnaire: its development and psychometric properties. *Spine (Phila Pa 1976)* 27:2116–2123. doi:[10.1097/01.brs.0000025729.35559.28](https://doi.org/10.1097/01.brs.0000025729.35559.28) (discussion 2124)
64. Bolton JE (2004) Sensitivity and specificity of outcome measures in patients with neck pain: detecting clinically significant improvement. *Spine (Phila Pa 1976)* 29:2410–2417 (discussion 2418)
65. Cleland JA, Fritz JM, Whitman JM, Palmer JA (2006) The reliability and construct validity of the neck disability index and patient specific functional scale in patients with cervical radiculopathy. *Spine (Phila Pa 1976)* 31:598–602. doi:[10.1097/01.brs.0000201241.90914.22](https://doi.org/10.1097/01.brs.0000201241.90914.22)
66. Fehlings MG WJ, Kopjar B, Yoon S, Arnold P, Massicotte EMM, Vaccaro A, Brodke DS, Shaffrey C, Smith JS, Woodward E, Banco RJ, Chapman J, Janssen M, Bono C, Sasso R, Dekutoski M, Gokaslan ZL (2013) Efficacy and safety of surgical decompression in patients with cervical spondylotic myelopathy: results

- of the AOSpine North America multi-centre study. *J Bone Joint Surg* (in press)
67. Singh A, Gnanalingham KK, Casey AT, Crockard A (2005) Use of quantitative assessment scales in cervical spondylotic myelopathy—survey of clinician'' attitudes. *Acta Neurochir (Wien)* 147:1235–1238. doi:[10.1007/s00701-005-0639-7](https://doi.org/10.1007/s00701-005-0639-7)
68. Tetreault LA, Karpova A, Fehlings MG (2013) Predictors of outcome in patients with degenerative cervical spondylotic myelopathy undergoing surgical treatment: results of a systematic review. *Eur Spine J*. doi:[10.1007/s00586-013-2658-z](https://doi.org/10.1007/s00586-013-2658-z)
69. Nakashima H, Yukawa Y, Ito K, Machino M, Kanbara S, Morita D, Takahashi H, Imagama S, Ito Z, Ishiguro N, Kato F (2012) Prediction of lower limb functional recovery after laminoplasty for cervical myelopathy: focusing on the 10-s step test. *Eur Spine J* 21:1389–1395. doi:[10.1007/s00586-012-2241-z](https://doi.org/10.1007/s00586-012-2241-z)
70. King JT Jr, Moossy JJ, Tsevat J, Roberts MS (2005) Multimodal assessment after surgery for cervical spondylotic myelopathy. *J Neurosurg Spine* 2:526–534