SUPPLEMENTARY INFORMATION

Supplementary Table 1. Full search phrases used for MEDLINE and EMBASE on 4 August 2020

	Ovid MEDLINE	**2389 articles**
Cervical spine concept		
1	exp Cervical Vertebrae/ or exp Cervical Cord/ or cervical.tw	
DCM concept		
2	Exp Spinal Cord Diseases/ or Exp Spinal Diseases/	
3	degenerat*.tw	
4	2 and 3	
5	Myelopath*.tw	
6	Myeloradiculopath*.tw	
7	Radiculopath*.tw	
8	Exp Spinal Cord Compression/	
9	Exp “Ossification of the Posterior Longitudinal Ligament”/	
10	Ossification of the Posterior Longitudinal Ligament.tw	
11	OPLL.tw	
12	Exp Spinal Stenosis/	
13	Cervical stenosis.tw	
14	Exp Spondylosis/	
15	Spondylosis.tw	
16	Spondylotic.tw	
17	Degenerative cervical myelopathy.tw	
18	DCM.tw	
19	Cervical spondylotic myelopathy.tw	
20	CSM.tw	
21	4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20	
22	1 and 21	
Tools for function concept (neurological/gait/mobility centric) concept		
23	Japanese Orthopaedic Association score.tw	
24	JOA.tw	
25	modified JOA.tw	
26	mJOA.tw	
27	Graded Redefined Assessment of Sensation Strength and Prehension.tw	
28	GRASSP.tw	
29	Quick Disability of the Arm Shoulder and Hand.tw	
30	QuickDASH.tw	

31 Myelopathy Disability Index.tw
32 MDI.tw
33 Nurick score.tw
34 Neck functional disability scale.tw
35 NFDS.tw
36 Neck Disability Index.tw
37 NDI.tw
38 Cooper myelopathy scale.tw
39 CMS.tw
40 European myelopathy score.tw
41 EMS.tw
42 Bournemouth questionnaire.tw
43 BQ.tw
44 Cervical spine outcomes questionnaire.tw
45 CSOQ.tw
46 Patient specific functional scale.tw
47 PSFS.tw
48 World Health Organization Quality of Life Instruments.tw
49 WHOQOL.tw
50 Grip and release test.tw
51 GRT.tw
52 Grip Dynamometer.tw
53 Triangle step test.tw
54 Foot tapping test.tw
55 30 m walking test.tw
56 30MWT.tw
57 10 m walking test.tw
58 10MWT.tw
59 Berg Balance Scale.tw
60 BBS.tw
61 GAITRite.tw
62 10 second step test.tw
63 9 hole peg test.tw
64 Prolo.tw
65 Mental component score.tw
66 MCS.tw
67 Physical component score.tw
68 PCS.tw

69	Hospital anxiety depression scale.tw
70	HADS.tw
71	Global rating of change.tw
72	GROC.tw
73	23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72
Tools for QOL concept (including pain) concept	
74	Exp “Quality of Life”/ or exp “Surveys and Questionnaires”/
75	Short Form Health Survey.tw
76	SF-36.tw
77	SF-12.tw
78	EQ-5D.tw
79	Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire.tw
80	JOACMEQ.tw
81	Visual Analogue Scale.tw
82	VAS.tw
83	Likert scale.tw
84	Numeric pain rating scale.tw
85	NPRS.tw
86	North American Spine Satisfaction.tw
87	NASS.tw
88	74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84 or 85 or 86 or 87
Psychometric concept	
89	Exp Psychometrics/
90	Pyschometr*.tw
91	(clinimetr* or clinometr*).tw.
92	Outcome assessment*.tw
93	exp Health Status Indicators/
94	Exp “Reproducibility of Results”/
95	Reproducib*.tw
96	Exp Validation Study/
97	exp Discriminant Analysis/
98	(reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or internal consistency).tw
99	(cronbach* and (alpha or alphas)).tw.
100	(item and (correlation* or selection* or reduction*)).tw

101 (agreement or precision or imprecision or precise values or test–retest).tw
102 (reliab* and (test or retest)).tw

103 (stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intraobserver or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or interassay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappas or repeatab*).tw
104 ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).tw
105 (generaliza* or generalisa* or concordance).tw
106 (intraclass and correlation*).tw
107 Exp Observer Variation/
108 Observer variation.tw
109 (multitrait and scaling and (analysis or analyses)).tw
110 Measurement error*.tw
111 (item discriminant or interscale correlation* or error or errors or individual variability).tw
112 (variability and (analysis or values)).tw
113 (uncertainty and (measurement or measuring)).tw
114 Exp Diagnostic Errors/
115 Exp Data accuracy/
116 Exp Dimensional Measurement Accuracy/
117 Accuracy.tw
118 ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).tw
119 Minimally clinically important difference*.tw
120 MCID.tw
121 (small* and (real or detectable) and (change or difference)).tw
122 (meaningful change or ceiling effect or floor effect or Item response model or IRT or Rasch or Differential item functioning or DIF or computer adaptive testing or item bank or cross-cultural equivalence).tw
123 Exp Bias/ or exp Selection Bias/
124 Bias.tw
125 Exp “Predictive Value of Test”/

126	89 or 90 or 91 or 92 or 93 or 94 or 95 or 96 or 97 or 98 or 99 or 100 or 101 or 102 or 103 or 104 or 105 or 106 or 107 or 108 or 109 or 110 or 111 or 112 or 113 or 114 or 115 or 116 or 117 or 118 or 119 or 120 or 121 or 122 or 123 or 124 or 125
Combined concepts	
127	73 or 88
128	22 and 126 and 127

EMBASE **1550 articles**

Cervical spine concept	
1	exp Cervical Vertebra/ or cervical spine/ or exp Cervical spinal cord/ or cervical.tw
DCM concept	
2	Exp Spinal Cord Disease/ or Exp Spine Disease/
3	Exp degeneration/
4	degenerat*.tw
5	3 or 4
6	2 and 5
7	Myelopath*.tw
8	Myeloradiculopath*.tw
9	Exp radiculopathy/
10	Radiculopath*.tw
11	Exp Spinal Cord Compression/
12	Exp Posterior Longitudinal Ligament/ and exp ossification/
13	Ossification of the Posterior Longitudinal Ligament.tw
14	OPLL.tw
15	Exp vertebral canal stenosis/
16	Cervical stenosis.tw
17	Exp Cervical Spondylosis/
18	Exp Spondylosis/
19	Spondylosis.tw
20	Spondylotic.tw
21	Exp Cervical myelopathy/
22	Degenerative cervical myelopathy.tw
23	DCM.tw
24	Exp Cervical spondylotic myelopathy/
25	Cervical spondylotic myelopathy.tw
26	CSM.tw
27	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26

28	1 and 27
Tools for function concept (neurological/gait/mobility centric) concept	
29	Exp Japanese Orthopaedic Association score/
30	Japanese Orthopaedic Association score.tw
31	JOA.tw
32	modified JOA.tw
33	mJOA.tw
34	Exp “Disabilities of the Arm, Shoulder and Hand (score)”/
35	Graded Redefined Assessment of Sensation Strength and Prehension.tw
36	GRASSP.tw
37	Quick Disability of the Arm Shoulder and Hand.tw
38	QuickDASH.tw
39	Myelopathy Disability Index.tw
40	MDI.tw
41	Exp “Nurick (grade)”/
42	Nurick score.tw
43	Neck functional disability scale.tw
44	NFDS.tw
45	Exp Neck Disability Index/
46	Neck Disability Index.tw
47	NDI.tw
48	Cooper myelopathy scale.tw
49	CMS.tw
50	European myelopathy score.tw
51	EMS.tw
52	Bournemouth questionnaire.tw
53	BQ.tw
54	Cervical spine outcomes questionnaire.tw
55	CSOQ.tw
56	Patient specific functional scale.tw
57	PSFS.tw
58	World Health Organization Quality of Life Instruments.tw
59	WHOQOL.tw
60	Grip and release test.tw
61	GRT.tw
62	Grip Dynamometer.tw
63	Triangle step test.tw
64	Foot tapping test.tw

65	30 m walking test.tw
66	30MWT.tw
67	10 m walking test.tw
68	10MWT.tw
69	Berg Balance Scale.tw
70	BBS.tw
71	GAITRite.tw
72	10 second step test.tw
73	9 hole peg test.tw
74	Prolo.tw
75	Mental component score.tw
76	MCS.tw
77	Physical component score.tw
78	PCS.tw
79	Hospital anxiety depression scale.tw
80	HADS.tw
81	Global rating of change.tw
82	GROC.tw
83	29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82
Tools for QOL concept (including pain) concept	
84	Short Form Health Survey.tw
85	Exp Short Form 36/
86	SF-36.tw
87	Exp Short Form 12/
88	SF-12.tw
89	Exp “European Quality of Life 5 Dimensions questionnaire”/
90	EQ-5D.tw
91	Exp Japanese Orthopaedic Association Cervical Myelopathy Evaluation/
92	Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire.tw
93	JOACMEQ.tw
94	Exp visual analog scale/
95	Visual Analogue Scale.tw
96	VAS.tw
97	Exp Likert scale/
98	Likert scale.tw

99	Numeric pain rating scale.tw
100	NPRS.tw
101	North American Spine Satisfaction.tw
102	NASS.tw
103	84 or 85 or 86 or 87 or 88 or 89 or 90 or 91 or 92 or 93 or 94 or 95 or 96 or 97 or 98 or 99 or 100 or 101 or 102
Psychometric concept	
104	Exp Psychometry/
105	Pyschometr*.tw
106	(clinimetr* or clinometr*).tw.
107	Outcome assessment*.tw
108	exp Health Status Indicator/
109	Exp Reproducibility/
110	Reproducib*.tw
111	Exp Validation Study/
112	exp Discriminant Analysis/
113	(reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or internal consistency).tw
114	(cronbach* and (alpha or alphas)).tw.
115	(item and (correlation* or selection* or reduction*)).tw
116	(agreement or precision or imprecision or precise values or test–retest).tw
117	(reliab* and (test or retest)).tw
118	(stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intraobserver or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or interassay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappas or repeatab*).tw
119	((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).tw
120	(generaliza* or generalisa* or concordance).tw
121	(intraclass and correlation*).tw
122	Exp Observer Variation/
123	Observer variation.tw
124	(multitrait and scaling and (analysis or analyses)).tw
125	Measurement error*.tw
126	(item discriminant or interscale correlation* or error or errors or individual variability).tw

127	(variability and (analysis or values)).tw
128	(uncertainty and (measurement or measuring)).tw
129	Exp Diagnostic Error/
130	Exp Data accuracy/
131	Exp Dimensional Measurement Accuracy/
132	((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).tw
133	Minimally clinically important difference*.tw
134	MCID.tw
135	(small* and (real or detectable) and (change or difference)).tw
136	(meaningful change or ceiling effect or floor effect or Item response model or IRT or Rasch or Differential item functioning or DIF or computer adaptive testing or item bank or cross-cultural equivalence).tw
137	Exp Bias/ or exp Selection Bias/
138	Bias.tw
139	104 or 105 or 106 or 107 or 108 or 109 or 110 or 111 or 112 or 113 or 114 or 115 or 116 or 117 or 118 or 119 or 120 or 121 or 122 or 123 or 124 or 125 or 126 or 127 or 128 or 129 or 130 or 131 or 132 or 133 or 134 or 135 or 136 or 137 or 138
Combined concepts	
140	83 or 103
141	28 and 139 and 140

Supplementary Table 2. Study characteristics

Study	Country	Sample size	Psychometric properties	Outcome measures
Auffinger, Lall (1)	United States	30	MCID/SCB Measurement error	NDI VAS for pain SF-36
Augusto, Diniz (2)	Brazil	30	Cross-cultural validity/Measurement invariance Reliability Responsiveness	JOA NDI
Azimi, Rezaei (3)	Iran	87	Cross-cultural validity/Measurement invariance Responsiveness	JOACMEQ
Badhiwala, Witiw (4)	Canada	606	MCID/SCB	SF-36 mJOA Nurick Scale NDI
Bohm, Fehlings (5)	Multicenter/ Global	601	Reliability Hypotheses testing for construct validity Responsiveness	Walking tests (timed or steps) mJOA Nurick Scale NDI SF-36
Carreon, Glassman (6)	United States	505	MCID/SCB	NDI SF-36 "Numeric rating scale" for pain
Chang, Kong (7)	Korea	108	Reliability	CT / CTM
Chiba, Kato (8)	Japan		Reliability	X-rays
Chien, Lai (9)	Taiwan	45	Responsiveness MCID/SCB	JOACMEQ NDI
Chiu and Pang (10)	Hong Kong	72	Internal consistency Reliability Content validity Hypotheses testing for construct validity Criterion validity Measurement error	BBS mJOA

			Responsiveness	
Fukui, Chiba (11)	Japan	368	Content validity	JOACMEQ
Fukui, Chiba (12)	Japan	201	Reliability	JOACMEQ
Goyal, Murphy (13)	United States	118	Responsiveness	NDI SF-12
Gwinn, Iannotti (14)	United States	20	Reliability	X-rays Cobb's method
Hosono, Sakaura (15)	Japan	30	Reliability Criterion validity	Grip-and-release test JOA
Hosono, Takenaka (16)	Japan	48	Responsiveness	Grip-and-release test JOA
Kang, Lee (17)	Korea	82	Reliability	MRI (not DTI)
Kato, Oshima (18)	Japan	92	Measurement error Hypotheses testing for construct validity Responsiveness	JOA mJOA JOACMEQ NDI SF-12
Kato, Oshima (19)	Japan	101	Measurement error Criterion validity MCID/SCB	JOA Likert scale
Kato, Oshima (20)	Japan	101	Measurement error Criterion validity MCID/SCB	JOACMEQ NDI EQ-5D SF-36 Likert scale
King and Roberts (21)	United States	88	Internal consistency	SF-36
Ko, Choi (22)	Korea	357	Reliability	MRI (not DTI)
Kopjar, Tetreault (23)	USA	277	Responsiveness Hypotheses testing for construct validity Internal consistency	mJOA Nurick Scale NDI SF-36 Walking tests (timed or steps)
Latimer, Haden (24)	England	70	Responsiveness	SF-36 NDI

				VAS for pain MDI
Longo, Berton (25)	Italy	75	Cross-cultural validity/Measurement invariance Reliability Internal consistency Hypotheses testing for construct validity Responsiveness Criterion validity	mJOA Nurick Scale NDI SF-36
Lubelski, Alvin (26)	United States	119	Hypotheses testing for construct validity Responsiveness Criterion validity	mJOA Nurick Scale EQ-5D
Mihara, Kondo (27)	Japan	270	Hypotheses testing for construct validity	Grip-and-release test Triangle step test
Nakamoto, Oshima (28)	Japan	94	Internal consistency Hypotheses testing for construct validity Criterion validity	QuickDASH JOA NDI SF-36 "Numeric rating scale" for pain
Nakashima, Yukawa (29)	Japan	101	Hypotheses testing for construct validity	
Nicholson, Millhouse (30)	United States	235	Hypotheses testing for construct validity	MRI (not DTI) mJOA NDI SF-12 "Numeric rating scale" for pain Isihara's Cervical Curvature Index
Nikaido, Kikuchi (31)	Japan	87	Hypotheses testing for construct validity)	JOACMEQ SF-36
Numasawa, Ono (32)	Japan	126	Hypotheses testing for construct validity Responsiveness Reliability	JOA Foot tapping test Grip-and-release test
Olindo, Signate (33)	France	40	Reliability	9-Hole peg test MRI (not DTI) Nurick Scale mJOA Walking tests (timed or steps)

Park, Kim (34)	Korea	100	Reliability	MRI (not DTI)
Pratali, Smith (35)	Brazil		Cross-cultural validity	mJOA
Pratali, Smith (36)	Brazil	55	Reliability	mJOA
Rhee, Shi (37)	United States	100	Criterion validity Reliability Content validity	mJOA
Sato, Horikoshi (38)	Japan	66	Hypotheses testing for construct validity	MRI (DTI) JOA
Shim, Lee (39)	Korea	79	Reliability Criterion validity	MRI (not DTI)
Singh and Crockard (40)	England	100	Internal consistency Responsiveness	Odom's Criteria Nurick Scale Ranawat classification of disease severity MDI JOA EMS SF-36
Singh and Crockard (41)	United Kingdom	41	Hypotheses testing for construct validity	Walking tests (timed or steps) MDI Nurick Scale
Singh, Gnanalingham (42)	England	105	Internal consistency Criterion validity Responsiveness	SF-12 SF-36
Spurgas, Abbas (43)	USA	35	MCID/SCB	VAS for pain NDI SF-12 mJOA
Tetreault, Nouri (44)	Canada	755	MCID/SCB	mJOA NDI
Thakar and Rajshekhar (45)	India	51	MCID/SCB Responsiveness	VAS for pain Nurick Scale SF-36
Thakar, Christopher (46)	India	70	Internal consistency Criterion validity	WHOQOL-Bref SF-36

			Responsiveness MCID/SCB	Nurick Scale
Wada, Fukui (47)	Japan	137	Responsiveness	JOACMEQ JOA 10-s step test
Witayakom, Paholpak (48)	Thailand	70	Cross-cultural validity/Measurement invariance Reliability Internal consistency Hypotheses testing for construct validity	JOACMEQ SF-36
Yonenobu, Abumi (49)	Japan	29	Reliability	JOA
Yukawa, Kato (50)	Japan	163	Hypotheses testing for construct validity Reliability Criterion validity	10-s step test JOA Grip-and-release test
Zhang, Zhou (51)	China	142	Internal consistency Responsiveness MCID/SCB	SF-36 mJOA
Zhou, Zhang (52)	China	113	MCID/SCB Measurement error	mJOA SF-36

Supplementary Table 3. Interpretability (i.e., MCID and SCB).

Instrument		Result summary	Overall rating
EQ-5D		MCID: 0.05; total sample size: 101	Sufficient
JOA		MCID: 2.5; total sample size: 101	Sufficient
JOACMEQ			
	Bladder function	MCID: 6.0; total sample size: 78	Sufficient
	Cervical spine function	MCID: 2.5; total sample size: 179	Sufficient
	Lower extremity function	MCID range 2.5–9.4; total sample size: 179	Sufficient
	QOL	MCID range 8.5–9.5; total sample size: 179	Sufficient
	Upper extremity function	MCID range 2.5–13.0; total sample size: 179	Sufficient
mJOA		MCID range 1.3–3.1; total sample size: 868	Sufficient
		SCB: 14; total sample size: 35	Indeterminate
NDI		MCID range 5–13; total sample size: 108	Sufficient
		SCB range 9.5–36; total sample size: 65	Indeterminate
Pain, "Numeric rating scale" (Arm pain)		MCID: 2.5; total sample size: 30 SCB: 3.5; total sample size: 30	Indeterminate
Pain, "Numeric rating scale" (Neck pain)		MCID: 2.5; total sample size: 30 SCB: 3.5; total sample size: 30	Indeterminate
SF-12			

	MCS	SCB: 51.5; total sample size: 35	Indeterminate
	PCS	SCB: 30.1; total sample size: 35	Indeterminate
SF-36			
	MCS	MCID range 3.0–7.4; total sample size: 749	Sufficient
	PCS	"MCID range 3.9–9.6; total sample size: 890 SCB: 16; total sample size: 30"	Sufficient
VAS for pain		MCID range 0.4–2.7; total sample size: 30	Sufficient
		SCB: 1.1; total sample size: 30	Indeterminate

Supplementary Table 4. Feasibility assessment.

Tool	Time (min)	Equipment	Training	License	Money	Ease of administration	Overall assessment
10-s step test	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
30MWT	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
9-Hole peg test	5	Yes	No	No	No	Barriers	Barriers
Berg Balance Scale	>15	Yes	Yes	No	No	Barriers	Barriers
Cobb's method (C2-C7)	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
CT (Tsuyama's classification, 2D & 3D)	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
CT (Tsuyama's classification, lateral + axial)	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
EQ-5D	5	Minimal	No	Yes	Yes	Minimal barriers	Minimal barriers
European Myelopathy Scale	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
Foot tapping test	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
Grip-and-release test	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
Isihara's Cervical Curvature Index	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
JOA	5	No	No	No	No	No barriers	No barriers

JOACMEQ	5–15	Minimal	No	No	No	Minimal barriers	Minimal barriers
MDI	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
mJOA	5	No	No	No	No	No barriers	No barriers
MRI (Depiction of intramedullary hyperintensity at eight cervical disc levels, T2W, 1.5-T or 3-T)	5–15	Minimal	No	No	No	Minimal barriers	Minimal barriers
MRI (Kang's classification, 1.5-T or 3-T)	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
MRI (Muhle's classification, 1.5-T)	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
MRI (Vaccaro's classification, 1.5-T)	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
NDI	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
Nurick scale	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
P-mJOA	5	Minimal	No	No	No	Minimal barriers	Minimal barriers

Arm pain score	5	No	No	No	No	No barriers	No barriers
Neck pain score	5	No	No	No	No	No barriers	No barriers
QuickDASH	5	Minimal	No	Yes	Yes	Minimal barriers	Barriers
Ranawat classification of disease severity	5	No	No	No	No	No barriers	No barriers
SF-12	5	Minimal	No	Yes	Yes	Minimal barriers	Barriers
SF-36	5–15	Minimal	No	Yes	Yes	Minimal barriers	Barriers
Triangle step test	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
VAS for pain	5	Minimal	No	No	No	Minimal barriers	Minimal barriers
WHOQOL-Bref	5–15	Minimal	No	No	No	Minimal barriers	Minimal barriers
X-rays (Computer-assisted measurement of length & thickness)	5–15	Minimal	No	No	No	Minimal barriers	Minimal barriers

Supplementary Table 5. Content validity.

Instrument	Result summary	Overall rating	Quality of evidence
BBS	Patient comprehensibility: Item discrimination index >0.589	Indeterminate	Very low
JOACMEQ	Patient comprehensibility: "No questions elicited no answer or “I am not sure” in more than 5% of patients"	Indeterminate	Very low
P-mJOA	Patient comprehensibility: "In patients preferring to complete the mJOA them- selves, the most popular answers were: “ease of answering the questions” (n = 33), “understanding of the questions” (n = 17)"	Indeterminate	Very low

Supplementary Table 6. Internal consistency.

Instrument	Result summary	Overall rating	Quality of evidence
BBS	Cronbach's alpha range 0.95–0.98; consistent; total sample size: 72	Indeterminate	Moderate
European Myelopathy Scale	Cronbach's alpha: 0.68; consistent; total sample size: 100	Indeterminate	Low
JOA	Cronbach's alpha: 0.72; consistent; total sample size: 100	Indeterminate	Low
JOACMEQ	Cronbach's alpha: 0.91; total sample size: 70	Indeterminate	Moderate
Bladder function	Cronbach's alpha range 0.77–0.78; consistent; total sample size: 157	Indeterminate	High
Cervical spine function	Cronbach's alpha range 0.75–0.88; consistent; total sample size: 157	Indeterminate	High
QOL	Cronbach's alpha range 0.80–0.86; consistent; total sample size: 157	Indeterminate	High
Upper extremity function	Cronbach's alpha range 0.72–0.74; consistent; total sample size: 157	Indeterminate	High
MDI	Cronbach's alpha: 0.92; consistent; total sample size: 100	Indeterminate	Low
mJOA	Cronbach's alpha range 0.60–0.63; consistent; total sample size: 352	Indeterminate	High
QuickDASH	Cronbach's alpha: 0.94; consistent; total sample size: 94	Indeterminate	Very low
SF-12	Cronbach's alpha: 0.77; consistent; total sample size: 105	Indeterminate	n/a

SF-36	Cronbach's alpha range 0.79–0.93; consistent; total sample size: 473	Indeterminate	n/a
WHOQOL-Bref	Cronbach's alpha range 0.86–0.87; consistent; total sample size: 38	Indeterminate	n/a

n/a = No info available

Supplementary Table 7. Cross-cultural validity.

Instrument	Result summary	Overall rating	Quality of evidence
JOA	Forward-backward translation [Brazilian Portuguese] Comprehension rate: >81.2%	Indeterminate	Very low
JOACMEQ	Forward-backward translation [Persian and Thai] No info available	Indeterminate	Very low
mJOA	Forward-backward translation [Brazilian Portuguese and Italian] No info available	Indeterminate	Very low

Supplementary Table 8. Reliability.

Instrument	Result summary	Overall rating	Quality of evidence
10-s step test	Test–retest stability: Spearman's rank correlation: 0.89; total sample size: 163	Indeterminate	Low
30MWT	Test–retest stability: Pearson's correlation range 0.89–1.00; total sample size: 16	Indeterminate	Very low
9-Hole peg test	Intra-observer reliability: ICC range 0.97–0.98; consistent; total sample size: 41 Inter-observer reliability: ICC range 0.97–0.99; consistent; total sample size: 41	Sufficient	Very low
BBS*	Test–retest stability: ICC: 0.99; total sample size: 32 Inter-observer reliability: ICC: 0.99; total sample size: 32	Sufficient	Very low
	Test–retest stability: Kappa: 0.67; total sample size: 32 Inter-observer reliability: Kappa: 0.43; total sample size: 32	Insufficient	Very low
Cobb's method	Intra-observer reliability: ICC: 0.84; total sample size: 20 Inter-observer reliability: ICC: 0.77; total sample size: 20	Sufficient	Very low
CT	Intra-observer reliability:	Sufficient	Moderate

(Tsuyama's classification, 2D & 3D)	Kappa range 0.85–0.86; consistent; total sample size: 108 Inter-observer reliability: Kappa range 0.71–0.76; consistent; total sample size: 108		
CT (Tsuyama's classification, lateral + axial)	Intra-observer reliability: Kappa: 0.67; total sample size: 108 Inter-observer reliability: Kappa: 0.51; total sample size: 108	Insufficient	Moderate
Foot tapping test	Test–retest stability: Pearson's correlation range 0.90–0.93; total sample size: 126	Indeterminate	Low
Grip-and-release test	Inter-observer reliability: ICC: 0.99; total sample size: 30	Sufficient	Very low
JOA	Inter-observer reliability: ICC: 0.81; total sample size: 29	Sufficient	Very low
Bladder function	Intra-observer reliability: Kappa: 0.64; total sample size: 29 Inter-observer reliability: Kappa: 0.47; total sample size: 29	Insufficient	Very low
Motor function of fingers	Intra-observer reliability: Kappa: 0.68; total sample size: 29 Inter-observer reliability: Kappa: 0.53; total sample size: 29	Insufficient	Very low
Motor function of	Intra-observer reliability: Kappa: 0.50; total sample size: 29	Insufficient	Very low

	shoulder and elbow	Inter-observer reliability: Kappa: 0.31; total sample size: 29		
	Motor function of lower extremity	Intra-observer reliability: Kappa: 0.55; total sample size: 29 Inter-observer reliability: Kappa: 0.49; total sample size: 29	Insufficient	Very low
	Sensory function of lower extremity	Intra-observer reliability: Kappa: 0.44; total sample size: 29 Inter-observer reliability: Kappa: 0.34; total sample size: 29	Insufficient	Very low
	Sensory function of trunk	Intra-observer reliability: Kappa: 0.54; total sample size: 29 Inter-observer reliability: Kappa: 0.58; total sample size: 29	Insufficient	Very low
	Sensory function of upper extremity	Intra-observer reliability: Kappa: 0.51; total sample size: 29 Inter-observer reliability: Kappa: 0.42; total sample size: 29	Insufficient	Very low
JOACMEQ				
	Bladder function	Test–retest stability: ICC: 0.62; total sample size: 70	Insufficient	Very low
	Cervical spine function	Test–retest stability: ICC: 0.63; total sample size: 70	Insufficient	Very low
	Lower extremity function	Test–retest stability: ICC: 0.93; total sample size: 70	Sufficient	Very low
	QOL	Test–retest stability:	Sufficient	Very low

		ICC: 0.83; total sample size: 70		
	Upper extremity function	Test–retest stability: ICC: 0.93; total sample size: 70	Sufficient	Very low
mJOA		Test–retest stability: Spearman's rank correlation: 0.91; total sample size: 75	Indeterminate	Very low
		Intra-observer reliability: ICC: 0.87; total sample size: 55	Sufficient	Very low
		Inter-observer reliability: ICC: 0.97; total sample size: 55 Kappa: 0.80; total sample size: 75	Sufficient	Low
	Motor dysfunction of lower extremities	Inter-observer reliability: ICC: 0.73; total sample size: 75	Sufficient	Low
	Motor dysfunction of upper extremities	Inter-observer reliability: ICC: 0.77; total sample size: 75	Sufficient	Low
	Sensory dysfunction of sphincter dysfunction	Inter-observer reliability: ICC: 0.78; total sample size: 75	Sufficient	Low
	Sensory dysfunction of upper extremities	Inter-observer reliability: ICC: 0.93; total sample size: 75	Sufficient	Low
MRI (Depiction of intramedullary hyperintensity at eight cervical disc		Inter-observer reliability: Kendall's W range 0.72–0.78; total sample size: 79	Indeterminate	Very low

levels, T2W, 1.5-T or 3-T)			
MRI (Kang's classification, 1.5-T or 3-T)	Intra-observer reliability: Kappa: 0.67; total sample size: 439 ICC: 0.77, total sample size: 82 Inter-observer reliability: Kappa range 0.60–0.93; total sample size: 539 ICC range 0.74–0.75; total sample size: 82	Inconsistent	n/a
MRI (Muhle's classification, 1.5-T)	Intra-observer reliability: Kappa: 0.72; total sample size: 357 Inter-observer reliability: Kappa range 0.61; total sample size: 357	Inconsistent	n/a
MRI (Vaccaro's classification, 1.5-T)	Intra-observer reliability: Kappa: 0.71; total sample size: 357 Inter-observer reliability: Kappa range 0.69; total sample size: 357	Sufficient	Moderate
P-mJOA			
Motor dysfunction of lower extremities	Inter-observer reliability: Kappa: 0.61; total sample size: 755	Insufficient	Moderate
Motor dysfunction of upper extremities	Inter-observer reliability: Kappa: 0.66; total sample size: 755	Insufficient	Moderate
Sensory dysfunction of sphincter dysfunction	Inter-observer reliability: Kappa: 0.55; total sample size: 755	Insufficient	Moderate

Sensory dysfunction of upper extremities	Inter-observer reliability: Kappa: 0.55; total sample size: 755	Insufficient	Moderate
X-rays (Computer-assisted measurement of length)	Intra-observer reliability: ICC: 0.94; total sample size: 9 Inter-observer reliability: ICC: 0.93; total sample size: 9	Sufficient	Very low
X-rays (Computer-assisted measurement of thickness)	Intra-observer reliability: ICC: 0.96; total sample size: 9 Inter-observer reliability: ICC: 0.97; total sample size: 9	Sufficient	Very low

*Result ratings for BBS were split by statistic used due to their associated differences in sufficiency.

Supplementary Table 9. Measurement error.

Instrument	Result summary	Overall rating	Quality of evidence
BBS	MDC or SDC Distribution: 1.5; total sample size: 32	Indeterminate	n/a
EQ-5D	MDC or SDC Distribution: 0.13; total sample size: 101 Anchor: 0.04; total sample size: 101	Inconsistent	n/a
JOA	MDC or SDC Distribution: 1.0; total sample size: 101 Anchor: 2.5; total sample size: 101 LOA 1.2 (–1.2, 3.6); total sample size: 92	Sufficient	Very low
JOACMEQ			
Bladder function	MDC or SDC Distribution: 7.7; total sample size: 101	Insufficient	Very low
Cervical spine function	MDC or SDC Distribution: 12.9; total sample size: 101 Anchor: 12.5; total sample size: 101	Insufficient	Very low
Lower extremity function	MDC or SDC Distribution: 7.3; total sample size: 101 Anchor: 9.4; total sample size: 101	Inconsistent	n/a
QOL	MDC or SDC Distribution: 6.6; total sample size: 101 Anchor: 8.5; total sample size: 101	Sufficient	Very low

	Upper extremity function	MDC or SDC Distribution: 9.5; total sample size: 101 Anchor: 6.1; total sample size: 101	Sufficient	Very low
mJOA		MDC or SDC Distribution: 2.1; total sample size: 113	Inconsistent	Very low
		MCID range; total sample size: 868 Distribution: 1.2–1.4	Sufficient	High
NDI		MDC or SDC Distribution: 6.2%; total sample size: 101 Anchor: 5.2%; total sample size: 101	Insufficient	Very low
SF-36				
	MCS	MDC or SDC Distribution: 3.3–5.7; total sample size: 244	Inconsistent	n/a
		MCID; total sample size: 748 Distribution: 3.4–6.8	Inconsistent	n/a
	PCS	MDC or SDC Distribution: 5.2–5.7; total sample size: 214 Anchor: 4.9; total sample size: 101	Inconsistent	n/a
		MCID range; total sample size: 861 Distribution: 2.9–5.5 MCID; total sample size: 51 Distribution: 10	Inconsistent	n/a
VAS for pain		MDC or SDC Distribution: 3.1; total sample size: 30	Insufficient	Very low

	MCID range 24.0–30.0; total sample size: 51	Insufficient	Very low
WHOQOL-Bref			
PH	MCID Distribution: 8.2; total sample size: 38	Indeterminate	n/a
PS	MCID Distribution: 7.9; total sample size: 38	Indeterminate	n/a
SR	MCID Distribution: 8.0; total sample size: 38	Indeterminate	n/a
EN	MCID Distribution: 5.6; total sample size: 38	Indeterminate	n/a
PF	MCID Distribution: 10.5; total sample size: 38	Indeterminate	n/a
RP	MCID Distribution: 17.2; total sample size: 38	Indeterminate	n/a
BP	MCID Distribution: 13.2; total sample size: 38	Indeterminate	n/a
GH	MCID Distribution: 12.3; total sample size: 38	Indeterminate	n/a
VT	MCID Distribution: 10.8; total sample size: 38	Indeterminate	n/a
SF	MCID Distribution: 13.6; total sample size: 38	Indeterminate	n/a
RE	MCID Distribution: 18.0; total sample size: 38	Indeterminate	n/a
MH	MCID	Indeterminate	n/a

	Distribution: 11.2; total sample size: 38

n/a = No info available

Supplementary Table 10. Criterion validity.

Instrument	Result summary*	Overall rating	Quality of evidence
10-s step test	JOA Spearman's rank correlation: 0.66; total sample size: 163	Insufficient	High
BBS	mJOA AUC range 0.88–0.94; total sample size: 31	Sufficient	Low
Foot tapping test	JOA Pearson's correlation: 0.66; total sample size: 126 JOA MFLE Pearson's correlation: 0.70; total sample size: 126	Insufficient	High
Grip-and-release test	JOA Pearson's correlation: 0.72; total sample size: 30	Sufficient	Low
JOA	Likert scale, "Health transition question" AUC: 0.59; total sample size: 101 Likert scale, "Patient satisfaction question" AUC: 0.62; total sample size: 101	Insufficient	Very low
JOACMEQ			
Cervical spine function	Likert scale, "Health transition question" AUC: 0.58; total sample size: 101 Likert scale, "Patient satisfaction question"	Insufficient	Very low

		AUC: 0.58; total sample size: 101		
	Upper extremity function	Likert scale, "Health transition question" AUC: 0.66; total sample size: 101 Likert scale, "Patient satisfaction question" AUC: 0.65; total sample size: 101	Insufficient	Very low
	Lower extremity function	Likert scale, "Health transition question" AUC: 0.61; total sample size: 101 Likert scale, "Patient satisfaction question" AUC: 0.66; total sample size: 101	Insufficient	Very low
	QOL	Likert scale, "Health transition question" AUC: 0.70; total sample size: 101 Likert scale, "Patient satisfaction question" AUC: 0.66; total sample size: 101	Insufficient	Very low
mJOA		Nurick scale [convergent] Spearman's rank correlation: –0.41; total sample size: 119 Pearson's correlation range: –0.62 to –0.63; total sample size: 352	Sufficient	High
	Motor dysfunction of upper extremities	Nurick scale [convergent] Pearson's correlation range –0.42 to –0.42; total sample size: 352	Insufficient	High

	Motor dysfunction of lower extremities	Nurick scale [convergent] Pearson's correlation: –0.65 to –0.68; total sample size: 352	Sufficient	High
	Sensory dysfunction of upper extremities	Nurick scale [convergent] Pearson's correlation: –0.23; total sample size: 277	Insufficient	High
	Sensory dysfunction of sphincter dysfunction	Nurick scale [convergent] Pearson's correlation: –0.25; total sample size: 277	Insufficient	High
NDI		Likert scale, "Health transition question" AUC: 0.66; total sample size: 101 Likert scale, "Patient satisfaction question" AUC: 0.75; total sample size: 101	Inconsistent	n/a
P-mJOA		mJOA Spearman's rank correlation: 0.83; total sample size: 755	Sufficient	High
QuickDASH		JOA MFSE Spearman's rank correlation: –0.50; total sample size: 94 JOA SFUE Spearman's rank correlation: –0.32; total sample size: 94	Insufficient	Moderate
SF-36				
	PCS	Likert scale, "Health transition question" AUC: 0.67; total sample size: 101	Insufficient	Very low

	Likert scale, "Patient satisfaction question" AUC: 0.69; total sample size: 101		
WHOQOL-Bref			
PH	SF-36 PCS Pearson's correlation: 0.51; total sample size: 38 SF-36 MCS Pearson's correlation: 0.30; total sample size: 38	Inconsistent	n/a
PS	SF-36 PCS Pearson's correlation: 0.34; total sample size: 38 SF-36 MCS Pearson's correlation: 0.23; total sample size: 38	Insufficient	Low
SR	SF-36 PCS Pearson's correlation: 0.35; total sample size: 38 SF-36 MCS Pearson's correlation: 0.28; total sample size: 38	Insufficient	Low
EN	SF-36 PCS Pearson's correlation: 0.05; total sample size: 38 SF-36 MCS Pearson's correlation: 0.03; total sample size: 38	Insufficient	Low

n/a = No info available

*Instruments listed are comparators

Supplementary Table 11. Construct validity.

Instrument	Result summary*	Overall rating	Quality of evidence
10-s step test	Grip-and-release test [convergent] Spearman's rank correlation: 0.53; total sample size: 163	Sufficient	Moderate
30MWT	mJOA [convergent] Pearson's correlation: –0.44; total sample size: 16 MDI [convergent] Spearman's rank correlation: 0.65; total sample size: 41 Nurick scale [convergent] Pearson's correlation: 0.50; total sample size: 16 Spearman's rank correlation: 0.61; total sample size: 41	Sufficient	Moderate
	NDI Pearson's correlation: 0.21; total sample size: 16	Sufficient	Low
	SF-36 PCS Pearson's correlation: –0.35; total sample size: 16	Sufficient	Low
	SF-36 MCS Pearson's correlation: –0.20; total sample size: 16	Sufficient	Low
BBS	mJOA [convergent] Spearman's rank correlation: 0.81; total sample size: 72	Sufficient	Moderate

EQ-5D	mJOA AUC: 0.68; total sample size: 119 Nurick scale AUC: 0.61; total sample size: 119	Insufficient	High
Foot tapping test	Grip-and-release test [convergent] Pearson's correlation: 0.58; total sample size: 126	Sufficient	High
Isihara's Cervical Curvature Index	mJOA Pearson's correlation: 0.04; total sample size: 235	Sufficient	High
	SF-12 PCS Pearson's correlation: 0.06; total sample size: 235 SF-12 MCS Pearson's correlation: 0.11; total sample size: 235	Sufficient	High
	Pain, "Numeric rating scale" (Arm pain score) Pearson's correlation: –0.28; total sample size: 235 Pain, "Numeric rating scale" (Neck pain scores) Pearson's correlation: –0.27; total sample size: 235 NDI Pearson's correlation: –0.10; total sample size: 235	Sufficient	High
JOA	mJOA [convergent]	Sufficient	Low

		Spearman's rank correlation: 0.87; total sample size: 92		
		JOACMEQ QOL [convergent] Spearman's rank correlation: 0.41; total sample size: 92	Sufficient	Low
		SF-12 PCS Spearman's rank correlation: 0.50; total sample size: 92	Sufficient	Low
		SF-12 MCS Spearman's rank correlation: –0.05; total sample size: 92	Sufficient	Low
		NDI Spearman's rank correlation range –0.50 to –0.76; total sample size: 122	Sufficient	Moderate
JOACMEQ				
	QOL	NDI Spearman's rank correlation: –0.66; total sample size: 92	Sufficient	Low
	QOL	SF-12 PCS Spearman's rank correlation: 0.29; total sample size: 92	Insufficient	Low
		SF-12 MCS Spearman's rank correlation: 0.40; total sample size: 92		
MDI		Nurick scale [convergent] Spearman's rank correlation: 0.66; total sample size: 41	Sufficient	Low

mJOA		30MWT [convergent] Pearson's correlation: –0.38; total sample size: 193	Insufficient	High
		JOACMEQ QOL [convergent] Spearman's rank correlation: 0.41; total sample size: 92	Insufficient	Low
		EQ-5D Spearman's rank correlation: 0.42; total sample size: 119	Insufficient	High
		SF-36 PCS Pearson's correlation range: 0.30–0.30; total sample size: 338 SF-12 PCS Spearman's rank correlation: 0.47; total sample size: 92	Sufficient	High
		SF-36 MCS Pearson's correlation: 0.25–0.25; total sample size: 338 SF-12 MCS Spearman's rank correlation: 0.03; total sample size: 92	Sufficient	High
		NDI Spearman's rank correlation: –0.51; total sample size: 92 Pearson's correlation rage –0.33 to –0.34; total sample size: 336	Sufficient	High
	Motor dysfunction	30MWT [convergent]	Insufficient	High

of lower extremities	Pearson's correlation: –0.43; total sample size: 193		
	SF-36 PCS Pearson's correlation range: 0.31–0.50; total sample size: 338	Sufficient	High
	SF-36 MCS Pearson's correlation: 0.21; total sample size: 268	Sufficient	High
	NDI Pearson's correlation: –0.31; total sample size: 261	Sufficient	High
Motor dysfunction of upper extremities	30MWT [convergent] Pearson's correlation: –0.21; total sample size: 193	Insufficient	High
	SF-36 PCS Pearson's correlation: 0.22; total sample size: 268	Insufficient	High
	SF-36 MCS Pearson's correlation: 0.20; total sample size: 268	Sufficient	High
	NDI Pearson's correlation: –0.24; total sample size: 261	Sufficient	High
Sensory dysfunction of sphincter dysfunction	30MWT [convergent] Pearson's correlation: –0.23; total sample size: 193	Insufficient	High
	SF-36 PCS	Sufficient	High

		Pearson's correlation: 0.06; total sample size: 268 SF-36 MCS Pearson's correlation: 0.08; total sample size: 268		
		NDI Pearson's correlation: –0.16; total sample size: 261	Sufficient	High
	Sensory dysfunction of upper extremities	30MWT [convergent] Pearson's correlation: –0.05; total sample size: 193	Insufficient	High
		SF-36 PCS Pearson's correlation: 0.19; total sample size: 268 SF-36 MCS Pearson's correlation: 0.19; total sample size: 268	Sufficient	High
		NDI Pearson's correlation: –0.23; total sample size: 261	Insufficient	High
NDI		mJOA Pearson's correlation: –0.36; total sample size: 235	Sufficient	High
		SF-12 PCS Pearson's correlation: –0.54; total sample size: 235 SF-12 MCS	Sufficient	High

	Pearson's correlation: –0.40; total sample size: 235		
	Pain, "Numeric rating scale" (Arm pain score) Pearson's correlation: 0.68; total sample size: 235 Pain, "Numeric rating scale" (Neck pain scores) Pearson's correlation: 0.64; total sample size: 235	Sufficient	High
Nurick scale	EQ-5D Spearman's rank correlation: –0.28; total sample size: 119	Sufficient	High
Pain, "Numeric rating scale" (Arm pain scores)	mJOA Pearson's correlation: –0.19; total sample size: 235	Sufficient	High
	Pain, “Numeric rating scale” (Neck pain score) [convergent] Pearson’s correlation: 0.72; total sample size: 235	Sufficient	High
Pain, "Numeric rating scale" (Neck pain scores)	mJOA Pearson's correlation: –0.07; total sample size: 235	Sufficient	High
QuickDASH	SF-36 Spearman's rank correlation: –0.75; total sample size: 94	Sufficient	Moderate
	NDI and Pain, "Numeric rating scale" [convergent]	Sufficient	Moderate

	Spearman's rank correlation range 0.69–0.83; total sample size: 94		
SF-12			
MCS	mJOA Pearson's correlation: 0.19; total sample size: 235	Sufficient	High
	Pain, "Numeric rating scale" (Arm pain score) Pearson's correlation: –0.23; total sample size: 235 Pain, "Numeric rating scale" (Neck pain score) Pearson's correlation: –0.28; total sample size: 235	Sufficient	High
	NDI Spearman's rank correlation: –0.17; total sample size: 92	Sufficient	Moderate
	SF-12 PCS Pearson's correlation: 0.01; total sample size: 235	Sufficient	High
PCS	mJOA Pearson's correlation: 0.43; total sample size: 235	Sufficient	High
	Pain, "Numeric rating scale" (Arm pain score) Pearson's correlation: –0.44; total sample size: 235 Pain, "Numeric rating scale" (Neck pain score)	Sufficient	High

	Pearson's correlation: –0.41; total sample size: 235		
	NDI Spearman's rank correlation: –0.49; total sample size: 92	Sufficient	Moderate
	SF-12 MCS Spearman's rank correlation: –0.29; total sample size: 92	Sufficient	Low
Triangle step test	Grip-and-release test [convergent] Spearman's rank correlation: 0.55; total sample size: 270	Sufficient	High

*Instruments listed are comparators

Supplementary Table 12. Responsiveness.

Instrument	Result summary*	Overall rating	Quality of evidence
30MWT	30MWT SRM: 0.3; total sample size: 484	Insufficient	High
BBS	mJOA Sensitivity range 77.4-80.0; total sample size: 31 Specificity range 87.8–92.9; total sample size: 31	Sufficient	Low
EQ-5D	EQ-5D Mean change score: 0.06; total sample size: 108	Indeterminate	High
European Myelopathy Scale	EMS Normalised change: 0.18; total sample size: 99	Indeterminate	Very low
Foot tapping test	Foot tapping test Mean change score: 6; total sample size: 6	Indeterminate	Very low
Grip-and-release test	Grip-and-release test Spearman's rank correlation: 0.69; total sample size: 48	Sufficient	Very low
	JOA Spearman's rank correlation: 0.32; total sample size: 48	Insufficient	Low
JOA	mJOA Spearman's rank correlation: 0.75; total sample size: 92	Sufficient	Very low
	JOA	Indeterminate	Very low

		Mean change score range 4.6; total sample size: 126 Normalised change: 0.21; total sample size: 99 JOA MFLE Mean change score range 0.6; total sample size: 126		
JOACMEQ				
	Bladder function	JOACMEQ BF AUC: 0.82; total sample size: 78	Sufficient	Moderate
		JOACMEQ BF Mean change score: 18.0; total sample size: 87	Indeterminate	Very low
		JOACMEQ BF Effect size: 0.33; total sample size: 78	Insufficient	Moderate
	Cervical spine function	JOACMEQ CF AUC: 0.72; total sample size: 78	Sufficient	Moderate
		JOACMEQ CF Mean change score: 25.8; total sample size: 87	Indeterminate	Very low
		JOACMEQ CF Effect size: 0.28; total sample size: 78	Insufficient	Moderate
	Lower extremity function	JOACMEQ LEF AUC: 0.75; total sample size: 78	Sufficient	Moderate
		JOACMEQ LEF Mean change score: 28.4; total sample size: 87	Indeterminate	Very low

		JOACMEQ LEF Effect size: 0.02; total sample size: 78	Insufficient	Moderate
	Upper extremity function	JOACMEQ UEF AUC: 0.74; total sample size: 78	Sufficient	Moderate
		JOACMEQ UEF Mean change score: 10.7; total sample size: 87	Indeterminate	Very low
		JOACMEQ UEF Effect size: 0.17; total sample size: 78	Insufficient	Moderate
	QOL	JOACMEQ QOL AUC: 0.83; total sample size: 78	Sufficient	Moderate
		JOACMEQ QOL Mean change score: 23.7; total sample size: 87	Indeterminate	Very low
		JOACMEQ QOL Effect size: 0.46; total sample size: 78	Insufficient	Moderate
MDI		MDI Normalised change: 0.52; total sample size: 99	Indeterminate	Very low
mJOA		mJOA Effect size range 0.87–1.0; total sample size: 352	Sufficient	High
		mJOA Normalised change: 1.47; total sample size: 42	Indeterminate	Very low
NDI		Anchor-based approach AUC: 0.66; total sample size: 78	Insufficient	Moderate

	Effect size: 0.44; total sample size: 78		
	NDI Mean change score: –15.8; total sample size: 118	Indeterminate	Very low
Nurick scale	Nurick scale Normalised change: 0.42; total sample size: 99 Mean change score range –0.76 to –1.3; total sample size: 93	Indeterminate	Very low
Ranawat classification of disease severity	Ranawat classification of disease severity Normalised change: 0.34; total sample size: 99	Indeterminate	Very low
SF-12			
PCS	SF-12 PCS Mean change score: 8.17; total sample size: 118	Indeterminate	Very low
SF-36	SF-36 Normalised change: 0.32; total sample size: 99	Indeterminate	Very low
PCS	SF-36 PCS Effect size range: 0.84; total sample size: 142	Sufficient	Low
	SF-36 PCS Sensitivity: 0.85; total sample size: 105	Sufficient	Moderate
MCS	SF-36 MCS Effect size range: 0.81; total sample size: 142	Sufficient	Low
	SF-36 MCS	Sufficient	Moderate

	Sensitivity: 0.67; total sample size: 105		
WHOQOL-Bref			
PH	WHOQOL-Bref PH Effect size: 0.68; total sample size: 38	Insufficient	Low
PS	WHOQOL-Bref PS Effect size: 0.39; total sample size: 38	Insufficient	Low
SR	WHOQOL-Bref SR Effect size: 0.03; total sample size: 38	Insufficient	Low
EN	WHOQOL-Bref EN Effect size: 0.45; total sample size: 38	Insufficient	Low

*Instruments listed are comparators

REFERENCES

1. Auffinger BM, Lall RR, Dahdaleh NS, Wong AP, Lam SK, Koski T, et al. Measuring surgical outcomes in cervical spondylotic myelopathy patients undergoing anterior cervical discectomy and fusion: assessment of minimum clinically important difference. PLoS One. 2013;8(6):e67408.
2. Augusto MT, Diniz JM, Dantas FLR, Oliveira MF, Rotta JM, Botelho RV. Development of the Portuguese Version of the Modified Japanese Orthopaedic Association Score: Cross-Cultural Adaptation, Reliability, Validity, and Responsiveness. World Neurosurg. 2018;116:e1092-e7.
3. Azimi P, Rezaei O, Montazeri A. An outcome measure of functionality and quality of life in patients with cervical myelopathy. Iran Red Crescent Med J. 2014;16(6):e8102.
4. Badhiwala JH, Witiw CD, Nassiri F, Akbar MA, Jaja B, Wilson JR, et al. Minimum Clinically Important Difference in SF-36 Scores for Use in Degenerative Cervical Myelopathy. Spine (Phila Pa 1976). 2018;43(21):E1260-e6.
5. Bohm PE, Fehlings MG, Kopjar B, Tetreault LA, Vaccaro AR, Anderson KK, et al. Psychometric properties of the 30-m walking test in patients with degenerative cervical myelopathy: results from two prospective multicenter cohort studies. Spine J. 2017;17(2):211-7.
6. Carreon LY, Glassman SD, Campbell MJ, Anderson PA. Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J. 2010;10(6):469-74.
7. Chang H, Kong CG, Won HY, Kim JH, Park JB. Inter- and intra-observer variability of a cervical OPLL classification using reconstructed CT images. Clin Orthop Surg. 2010;2(1):8-12.
8. Chiba K, Kato Y, Tsuzuki N, Nagata K, Toyama Y, Iwasaki M, et al. Computer-assisted measurement of the size of ossification in patients with ossification of the posterior longitudinal ligament in the cervical spine. J Orthop Sci. 2005;10(5):451-6.
9. Chien A, Lai DM, Cheng CH, Wang SF, Hsu WL, Wang JL. Responsiveness of the Chinese versions of the Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire and Neck Disability Index in postoperative patients with cervical spondylotic myelopathy. Spine (Phila Pa 1976). 2015;40(17):1315-21.
10. Chiu AY, Pang MY. Assessment of psychometric properties of various balance assessment tools in persons with cervical spondylotic myelopathy. journal of orthopaedic & sports physical therapy. 2017;47(9):673-82.
11. Fukui M, Chiba K, Kawakami M, Kikuchi S, Konno S, Miyamoto M, et al. Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire (JOACMEQ): Part 2. Endorsement of the alternative item. J Orthop Sci. 2007;12(3):241-8.
12. Fukui M, Chiba K, Kawakami M, Kikuchi S, Konno S, Miyamoto M, et al. Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire: part 3. Determination of reliability. J Orthop Sci. 2007;12(4):321-6.
13. Goyal DKC, Murphy HA, Hollern DA, Divi SN, Nicholson K, Stawicki C, et al. Is the Neck Disability Index an Appropriate Measure for Changes in Physical Function After Surgery for Cervical Spondylotic Myelopathy? Int J Spine Surg. 2020;14(1):53-8.
14. Gwinn DE, Iannotti CA, Benzel EC, Steinmetz MP. Effective lordosis: analysis of sagittal spinal canal alignment in cervical spondylotic myelopathy. J Neurosurg Spine. 2009;11(6):667-72.

15. Hosono N, Sakaura H, Mukai Y, Kaito T, Makino T, Yoshikawa H. A simple performance test for quantifying the severity of cervical myelopathy. J Bone Joint Surg Br. 2008;90(9):1210-3.
16. Hosono N, Takenaka S, Mukai Y, Makino T, Sakaura H, Miwa T, et al. Postoperative 24-hour result of 15-second grip-and-release test correlates with surgical outcome of cervical compression myelopathy. Spine (Phila Pa 1976). 2012;37(15):1283-7.
17. Kang Y, Lee JW, Koh YH, Hur S, Kim SJ, Chai JW, et al. New MRI grading system for the cervical canal stenosis. AJR Am J Roentgenol. 2011;197(1):W134-40.
18. Kato S, Oshima Y, Oka H, Chikuda H, Takeshita Y, Miyoshi K, et al. Comparison of the Japanese Orthopaedic Association (JOA) score and modified JOA (mJOA) score for the assessment of cervical myelopathy: a multicenter observational study. PloS one. 2015;10(4):e0123022.
19. Kato S, Oshima Y, Matsubayashi Y, Taniguchi Y, Tanaka S, Takeshita K. Minimum Clinically Important Difference and Patient Acceptable Symptom State of Japanese Orthopaedic Association Score in Degenerative Cervical Myelopathy Patients. Spine (Phila Pa 1976). 2019;44(10):691-7.
20. Kato S, Oshima Y, Matsubayashi Y, Taniguchi Y, Tanaka S, Takeshita K. Minimum clinically important difference in outcome scores among patients undergoing cervical laminoplasty. Eur Spine J. 2019;28(5):1234-41.
21. King JT, Jr., Roberts MS. Validity and reliability of the Short Form-36 in cervical spondylotic myelopathy. J Neurosurg. 2002;97(2 Suppl):180-5.
22. Ko S, Choi W, Chae S. Comparison of inter- and intra-observer reliability among the three classification systems for cervical spinal canal stenosis. Eur Spine J. 2017;26(9):2290-6.
23. Kopjar B, Tetreault L, Kalsi-Ryan S, Fehlings M. Psychometric properties of the modified Japanese Orthopaedic Association scale in patients with cervical spondylotic myelopathy. Spine (Phila Pa 1976). 2015;40(1):E23-8.
24. Latimer M, Haden N, Seeley HM, Laing RJ. Measurement of outcome in patients with cervical spondylotic myelopathy treated surgically. Br J Neurosurg. 2002;16(6):545-9.
25. Longo UG, Berton A, Denaro L, Salvatore G, Denaro V. Development of the Italian version of the modified Japanese orthopaedic association score (mJOA-IT): cross-cultural adaptation, reliability, validity and responsiveness. Eur Spine J. 2016;25(9):2952-7.
26. Lubelski D, Alvin MD, Nesterenko S, Sundar SJ, Thompson NR, Benzel EC, et al. Correlation of quality of life and functional outcome measures for cervical spondylotic myelopathy. Journal of Neurosurgery: Spine. 2016;24(3):483-9.
27. Mihara H, Kondo S, Murata A, Ishida K, Niimura T, Hachiya M. A new performance test for cervical myelopathy: the triangle step test. Spine (Phila Pa 1976). 2010;35(1):32-5.
28. Nakamoto H, Oshima Y, Takeshita K, Chikuda H, Ono T, Taniguchi Y, et al. Usefulness of QuickDASH in patients with cervical laminoplasty. J Orthop Sci. 2014;19(2):218-22.
29. Nakashima H, Yukawa Y, Ito K, Machino M, Kanbara S, Morita D, et al. 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. 2011;20(8):1318-22.
30. Nicholson KJ, Millhouse PW, Pflug E, Woods B, Schroeder GD, Anderson DG, et al. Cervical Sagittal Range of Motion as a Predictor of Symptom Severity in Cervical Spondylotic Myelopathy. Spine (Phila Pa 1976). 2018;43(13):883-9.
31. Nikaido T, Kikuchi S, Yabuki S, Otani K, Konno S. 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 (Phila Pa 1976). 2009;34(23):2568-72.
32. Numasawa T, Ono A, Wada K, Yamasaki Y, Yokoyama T, Aburakawa S, et al. Simple foot tapping test as a quantitative objective assessment of cervical myelopathy. Spine (Phila Pa 1976). 2012;37(2):108-13.
33. Olindo S, Signate A, Richech A, Cabre P, Catonne Y, Smadja D, et al. Quantitative assessment of hand disability by the Nine-Hole-Peg test (9-HPT) in cervical spondylotic myelopathy. Journal of Neurology, Neurosurgery & Psychiatry. 2008;79(8):965-7.
34. Park H-J, Kim SS, Chung E-C, Lee S-Y, Park N-H, Rho M-H, et al. Clinical correlation of a new practical MRI method for assessing cervical spinal canal compression. American Journal of Roentgenology. 2012;199(2):W197-W201.
35. Pratali RR, Smith JS, Motta RL, Martins SM, Motta MM, Rocha RD, et al. A Brazilian Portuguese cross-cultural adaptation of the modified JOA scale for myelopathy. Clinics. 2017;72:103-5.
36. Pratali RR, Smith JS, Rocha RD, Matos TD, Defino HLA, Herrero C. RELIABILITY OF A BRAZILIAN PORTUGUESE TRANSLATED AND CROSS-CULTURALLY ADAPTED VERSION OF THE MJOA SCALE. Acta Ortop Bras. 2018;26(5):335-7.
37. Rhee JM, Shi WJ, Cyriac M, Kim JY, Zhou F, Easley KA, et al. The P-mJOA: A Patient-derived, Self-reported Outcome Instrument for Evaluating Cervical Myelopathy: Comparison with the mJOA. Clin Spine Surg. 2018;31(2):E115-e20.
38. Sato T, Horikoshi T, Watanabe A, Uchida M, Ishigame K, Araki T, et al. Evaluation of cervical myelopathy using apparent diffusion coefficient measured by diffusion-weighted imaging. AJNR Am J Neuroradiol. 2012;33(2):388-92.
39. Shim E, Lee JW, Lee E, Kang Y, Kang HS, Kang WY, et al. Cervical spondylotic myelopathy: diagnostic performance of radiologists with varying levels of experience in comparing MR images acquired using field strengths of 1.5 and 3 Tesla. Acta Radiol. 2019;60(10):1314-20.
40. Singh A, Crockard HA. Comparison of seven different scales used to quantify severity of cervical spondylotic myelopathy and post-operative improvement. J Outcome Meas. 2001;5(1):798-818.
41. Singh A, Crockard HA. Quantitative assessment of cervical spondylotic myelopathy by a simple walking test. Lancet. 1999;354(9176):370-3.
42. Singh A, Gnanalingham K, Casey A, Crockard A. Quality of life assessment using the Short Form-12 (SF-12) questionnaire in patients with cervical spondylotic myelopathy: comparison with SF-36. Spine (Phila Pa 1976). 2006;31(6):639-43.
43. Spurgas MP, Abbas SF, Szewczyk BS, Yim B, Ata A, German JW. The effect of length of follow-up on substantial clinical benefit thresholds in patients undergoing surgery for cervical degenerative myelopathy. Journal of Clinical Neuroscience. 2019;62:88-93.
44. Tetreault L, Nouri A, Kopjar B, Côté P, Fehlings MG. The Minimum Clinically Important Difference of the Modified Japanese Orthopaedic Association Scale in Patients with Degenerative Cervical Myelopathy. Spine (Phila Pa 1976). 2015;40(21):1653-9.
45. Thakar S, Rajshekhar V. Evaluation of pain as a preference-based health status measure in patients with cervical spondylotic myelopathy undergoing central corpectomy. Acta Neurochir (Wien). 2012;154(2):335-40.
46. Thakar S, Christopher S, Rajshekhar V. 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. J Neurosurg Spine. 2009;11(4):402-12.
47. Wada E, Fukui M, Takahashi K, Takeuchi D, Hashizume H, Kanamori M, et al. Japanese orthopaedic association cervical myelopathy evaluation questionnaire (JOACMEQ): Part 5. Determination of responsiveness. Journal of Orthopaedic Science. 2019;24(1):57-61.
48. Witayakom W, Paholpak P, Jirarattanaphochai K, Kosuwon W, Sirichativapee W, Wisanuyotin T, et al. Validation of the reliability of the Thai version of the Japanese Orthopaedic Association Cervical Myelopathy Evaluation Questionnaire (JOACMEQ). Journal of Orthopaedic Science. 2016;21(2):124-7.
49. Yonenobu K, Abumi K, Nagata K, Taketomi E, Ueyama K. Interobserver and intraobserver reliability of the japanese orthopaedic association scoring system for evaluation of cervical compression myelopathy. Spine (Phila Pa 1976). 2001;26(17):1890-4; discussion 5.
50. Yukawa Y, Kato F, Ito K, Horie Y, Nakashima H, Masaaki M, et al. "Ten second step test" as a new quantifiable parameter of cervical myelopathy. Spine (Phila Pa 1976). 2009;34(1):82-6.
51. Zhang Y, Zhou F, Sun Y. Assessment of health-related quality of life using the SF-36 in Chinese cervical spondylotic myelopathy patients after surgery and its consistency with neurological function assessment: a cohort study. Health Qual Life Outcomes. 2015;13(1):1-7.
52. Zhou F, Zhang Y, Sun Y, Zhang F, Pan S, Liu Z. Assessment of the minimum clinically important difference in neurological function and quality of life after surgery in cervical spondylotic myelopathy patients: a prospective cohort study. Eur Spine J. 2015;24(12):2918-23.