

## Tracking patient-reported outcomes in spinal disorders

Nikhil R. Nayak, John Mitchell Coats, Kalil G. Abdullah, Sherman C. Stein, Neil R. Malhotra

Department of Neurosurgery, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA

E-mail: Nikhil R. Nayak - [nikhil.nayak@uphs.upenn.edu](mailto:nikhil.nayak@uphs.upenn.edu); John Mitchell Coats - [jcoats@mail.med.upenn.edu](mailto:jcoats@mail.med.upenn.edu); Kalil G. Abdullah - [kalil.abdullah@uphs.upenn.edu](mailto:kalil.abdullah@uphs.upenn.edu); Sherman C. Stein - [sherman.stein@uphs.upenn.edu](mailto:sherman.stein@uphs.upenn.edu); \*Neil R. Malhotra - [neil.malhotra@uphs.upenn.edu](mailto:neil.malhotra@uphs.upenn.edu)

\*Corresponding author

Received: 23 April 15 Accepted: 24 June 15 Published: 08 October 15

### Abstract

**Background:** Patient-reported outcome measures (PROMs) quantify health status from the patient's point of view. While the number of published outcomes studies grows each year, so too has the number of instruments being reported, leading to confusion on which instruments are appropriate to use for various spinal conditions.

**Methods:** A broad search was conducted to identify commonly used PROMs in patients undergoing spinal surgery. We searched PubMed for combinations of terms related to anatomic location and a measure of patient-reported outcome in the title or text. We supplemented the search using the "related articles" feature of PubMed and by manually searching the bibliographies of selected articles.

**Results:** Major categories of PROMs in spine surgery include health-related quality-of-life, pain, and disease-specific disability, for which several different instrument options were identified and detailed. The minimal clinically important difference varies between instruments and differentiates statistical significance from clinical significance. In addition, the accurate estimation of costs has become a challenging but intrinsically linked variable to outcomes as increased attention is paid to the relative value of surgical interventions.

**Conclusion:** While a number of PROMs are available for tracking outcomes in spine surgery, only a handful appear to be widely used. At least one instrument from each category should be measured pre- and post-operatively to quantify treatment effect. In addition, while the primary goal is to select the most appropriate instruments for the patient's condition, one should keep in mind sustainability of efforts with regard to patient and administrative burden.

**Key Words:** Comparative-effectiveness, cost-effectiveness, patient-reported outcomes, quality-of-life, spine

### Access this article online

**Website:**

[www.surgicalneurologyint.com](http://www.surgicalneurologyint.com)

**DOI:**

10.4103/2152-7806.166892

**Quick Response Code:**



## INTRODUCTION

Tracking surgical outcomes is vital for clinical progress and demonstrating the value of interventions. While "outcomes" can refer operative metrics (e.g., blood loss, operative time), radiographic parameters (e.g., canal decompression, fusion rates), or physician-assigned scales (e.g., American Spinal Injury Association grade), the patient's perspective is generally omitted. Patient-reported outcome measures (PROMs), on the other hand, aim

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

**For reprints contact:** [reprints@medknow.com](mailto:reprints@medknow.com)

**How to cite this article:** Nayak NR, Coats JM, Abdullah KG, Stein SC, Malhotra NR. Tracking patient-reported outcomes in spinal disorders. *Surg Neurol Int* 2015;6:S490-9.

<http://surgicalneurologyint.com/Tracking-patient-reported-outcomes-in-spinal-disorders/>

to quantify health status from the patient's viewpoint without interpretation of responses by the clinician.<sup>[73]</sup>

PROMs quantify treatment impact in three major categories: global health-related quality-of-life (HRQoL), pain, and disease-specific disability. These instruments facilitate a variety of studies, most commonly comparative- and cost-effectiveness research. In addition, PROMs have been used in process improvement, such as identifying mismatches in patient/provider perceptions of health status or assessing the appropriateness of referrals, and they may eventually be used to benchmark clinical centers relative to national averages.<sup>[54,78,82,87]</sup> PROMs may also be analyzed in conjunction with clinical parameters such as radiographic markers to preoperatively stratify patient selection and predict outcomes.<sup>[43,67]</sup>

The US government has made research involving PROMs a high priority with the passage of the Patient Protection and Affordable Care Act in 2010 and development of the Patient Centered Outcomes Research Institute.<sup>[59,74]</sup> In addition, the UK National Institute for Health and Care Excellence (NICE) has made a collection of PROMs standard practice in four common surgical procedures since 2009.<sup>[14]</sup> While the number of publications reporting PROMs grows each year, so too has the number of PROMs being used, which leads to confusion on what an instrument actually measures, and which instruments are appropriate to use for various conditions.

In this review, we aim to describe basic concepts associated with PROMs, the most commonly used instruments in spine surgery, and their application for tracking outcomes in a spine practice.

### Criteria for evaluating an instrument

For an instrument to become widely adopted, it must generally meet a number of quality control criteria. For brevity, in this review, we have simplified the statistics, broadly terms “psychometrics,” presented and their interpretations [Table 1].

Major evaluation criteria include validity, reliability, and responsiveness.<sup>[60]</sup> Validity assesses how accurately an instrument measures what it intends to measure. A common method of determining validity is by calculating floor and ceiling effects, that is, if  $\geq 15\%$  of respondents achieve the lowest or highest possible scores, validity is considered limited.<sup>[55]</sup> Reliability measures the reproducibility of an instrument's results. It can be evaluated with a variety of statistics, most commonly Cronbach's  $\alpha$ . A Cronbach's  $\alpha$  of  $\geq 0.70$  is a rule of thumb for acceptable reliability.<sup>[81]</sup> Finally, responsiveness represents a PROM's ability to detect change. The most common statistic used to evaluate responsiveness is the area under the receiver-operating characteristic curve (AUC). AUC values  $\geq 0.70$  are often considered to have acceptable responsiveness.<sup>[81]</sup>

**Table 1: Common abbreviations used throughout the review**

Abbreviation	Full-term	Definition
PROM	Patient-reported outcome measure	Instruments used to quantify health status that may be compared at different time points and between patients
HRQoL	Health-related quality-of-life	A global quality-of-life score representing major aspects of physical and mental health
MCID	Minimal clinically important difference	The minimum change in an outcome instrument score deemed to be clinically significant
QALY	Quality-adjusted life year	A measure of quality-of-life that is, used in healthcare economic studies, such as cost-effectiveness research

The minimal clinically important difference (MCID) provides a threshold for meaningful change in the value of a statistic and is unique to each instrument, disease state, and patient population being studied. It is calculated based on either sample data distribution (i.e., standard deviation, effect size), “anchor” methods wherein changes in scores are linked to a binary question on whether or not the patient considers himself improved, or by professional consensus.<sup>[19]</sup> MCIDs are important when interpreting the results of a study, particularly studies with larger sample sizes, as statistically significant differences may arise which are not clinically meaningful.

In the following sections, we detail the more commonly used PROMs in spine surgery [Tables 2-4]. This review is not exhaustive of all instruments employed in the literature, but rather presents those most likely to be encountered by the typical spine surgeon. These instruments have largely been deemed valid, reliable, and responsive, although we have noted perceived limitations when applicable.

## GENERIC HEALTH-RELATED QUALITY-OF-LIFE INSTRUMENTS

Generic HRQoL instruments quantify the overall physical, social, and mental well-being of a patient and apply to a range of diseases, not just spinal conditions [Table 1]. The “domains” within a generic HRQoL instrument refer to specific categories of questions, such as those focused on physical functioning or pain; thus, there is overlap with other categories of PROMs. The results of generic HRQoL instruments may be presented as a nominal score without interpretation of the relative weight of each question, or the results may be presented as “utility scores” tied to studies on societal preferences for different health states.

Utility scores generally range from 0 (death) to 1 (perfect health), although negative numbers are also possible and reflect health states considered worse than

death.<sup>[85]</sup> Utility scores may be reported in conjunction with time to derive quality-adjusted life years (QALYs), the fundamental currency of effectiveness research. QALYs are determined by the change in utility over time multiplied by the time period. It should be noted that the calculation of QALYs relies specifically on utility scores, not nominal HRQoL scores, pain scores, or disease-specific disability scores.

The most commonly used generic HRQoL measures in the spine surgery literature are the EuroQOL-5D (EQ-5D) and the Short Form (SF) instruments.

**Table 2: Psychometric properties in evaluating outcomes instruments**

Property	Description	Statistic used
Reliability	Consistency or reproducibility of an instrument	Cronbach's $\alpha$
Validity	Accurately measuring what it intends to measure	Ceiling/floor effects
Responsiveness	Ability to detect change	Area under the receiver-operating characteristic curve
Meaningful change	Threshold for determining clinically meaningful benefit	Minimal clinically important difference

### EuroQOL-5D

The EQ-5D is a preference-based generic HRQoL instrument developed in 1990 by the EuroQOL Group.<sup>[25]</sup> Utility scores are generated from five questions on mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Scoring of the EQ-5D has been conducted in multiple population samples (US, UK, Dutch, etc.). The minimum and maximum scores of the EQ-5D vary based on the population set being used. For example, UK population utilities range from -0.594 to 1, while US population utilities range from -0.109 to 1.

There are two versions of the EQ-5D, one with three options for each question (EQ-5D 3L) and one with five options (EQ-5D 5L). The EQ-5D 3L has been the more commonly used of the two and describes 243 possible health states, in addition to death and a comatose state. The second component of both versions of the EQ-5D is a visual analog scale (VAS) where responders are asked to rate their current overall health state as a single mark on a scale from 0 to 100.

Advantages of the EQ-5D include its brevity, applicability to a wide range of diseases, direct utility scoring, and common application in the spine surgery literature. In addition, it has been recommended by the

**Table 3: Commonly used generic HRQoL and pain instruments**

Name	Abbreviation	Category	Number of domains	Number of questions	Range of scores	Composite scores	Scoring	Notes
EuroQOL-5D <sup>†</sup>	EQ-5D	HRQoL	5	5+VAS	-0.594-1	-	Utility score	Values based on the UK valuation set
Short Form-36 <sup>†</sup>	SF-36	HRQoL	8	36	0-100	Physical, mental	Nominal score	For each domain or component summary
Short Form-6D <sup>†</sup>	SF-6D	HRQoL	6	11	0.296-1	Physical, mental	Utility score	Values based on the UK valuation set
Visual Analog Scale	VAS	Pain	-	1	0-100	-	Raw	
Numerical Rating Scale	NRS	Pain	-	1	1-10	-	Raw	
Mcgill Pain Questionnaire	MPQ	Pain	-	15+VAS+PPI	0-45, 0-100 (VAS), 0-5 (PPI)	Affective, sensory	Raw	

<sup>†</sup>Higher scores indicate better HRQoL. HRQoL: Health-related quality-of-life; VAS: Visual analog scale; PPI: Present pain index

**Table 4: Commonly used disease-specific instruments**

Name	Abbreviation	Spinal segment	Number of domains	Number of questions	Range of scores	Scoring	MCID
Neck Disability Index	NDI	Cervical	1	10	0-50 or 0-100%	Raw or percentage	2.41-13.39
Cervical Spine Outcomes Questionnaire	CSOQ	Cervical	6	35	0-100*	Transformed	13-24*
Modified Japanese Orthopedic Association Myelopathy Scale <sup>†</sup>	mJOA	Cervical	3	6	0-18	Raw	-
Myelopathy Disability Index	MDI	Cervical	5	10	0-30 or 0-100%	Raw or percentage	N/A
Oswestry Disability Index	ODI	Lumbar	1	10	0-50 or 0-100%	Raw or percentage	12.8-15
Roland-Morris Disability Questionnaire	RMDQ	Lumbar	1	24	0-24	Raw	1-8
Quebec Back Pain Disability Scale	QBPDS	Lumbar	6	20	0-100	Raw	15-19
Japanese Orthopedic Association Back Pain Evaluation Questionnaire <sup>†</sup>	JOABPEQ	Lumbar	5	25	0-100	Transformed	N/A

<sup>†</sup>Higher scores indicate better function, \*Each domain individually scored. N/A: Not available; MCID: Minimal clinically important difference

US Public Health Service Panel on Cost-effectiveness in Health and Medicine and the UK NICE.<sup>[21,88]</sup> Criticisms of the EQ-5D include it being too simplistic and that it may suffer from ceiling effects. For example, a study by Brazier *et al.* of nearly 2000 British respondents in the general population found nearly 50% of respondents scored at the ceiling.<sup>[10]</sup> In addition, within the spine literature, there has been a very wide range of reported MCID values depending on preoperative diagnosis, many of which are multiple times higher than the 0.074 MCID threshold first described by Walters and Brazier in 2005.<sup>[63,64,85]</sup>

### Short-forms

The SF instruments are based on the RAND Corporation's 1989 Medical Outcomes Study. The original survey, a 116-item questionnaire across eight domains on physical functioning, role limitations, health perceptions, vitality, social functioning, mental health, and health transition was consolidated to the more popular 36-item SF-36.<sup>[86]</sup> The results of the SF-36 are generally reported in terms of the eight separate domain scores and/or two summary scores (physical component summary [PCS], mental component summary), each of which ranges from 0 to 100. The preference-based SF-6D was subsequently developed to directly generate utility scores, which range from 0.296 to 1.<sup>[86]</sup> Because of overlapping questions, the SF-36 can be "mapped" to the SF-6D to also derive utility scores.

Advantages of the SF instruments include widespread use, comprehensive sets of questions, and the ability to generate utility scores via the SF-6D. A major disadvantage of the commonly employed SF-36 is that considerably more time and effort is required to complete it compared to the EQ-5D. In addition, most published SF-36 results are presented as domain or summary scores, which cannot reliably be compared to effectiveness research using utility scores. One study specific to cervical spine surgery found considerable floor effects,<sup>[5]</sup> and there has been some concern that the SF-6D overestimates utility scores, as the lowest score for the SF-6D is 0.296 utility versus negative utility scores for the EQ-5D.

The same 2005 study by Walters *et al.* on 11 various diseases identified an MCID of 0.041 for SF-6D utility values after an intervention.<sup>[85]</sup> A commonly used MCID for the SF-36 PCS in spine surgery is 5 points.<sup>[18]</sup>

Finally, it should be noted that although the EQ-5D and SF-6D both derive utility scores, the instruments often generate very different values for the same disease state and should not be used interchangeably.<sup>[77]</sup> Studies on nonspinal pathology, such as heart disease and osteoarthritis, have reported that utility values significantly vary between instruments.<sup>[71,83]</sup> Moreover, small variations in utility may lead to very large differences in the perceived cost-utility of a procedure,

which may have implications for health technology assessment and resource allocation.

## PAIN PATHOLOGY

### Visual analog scale

The VAS is a one-dimensional scale for rating current pain on a continuum.<sup>[62]</sup> It is typically presented as a 10 centimeter (cm) scale with millimeter (mm) markers, on which the patient can mark any point on the scale.<sup>[22]</sup> Patients are asked to rate their pain score based on the past 24 hours. The general interpretation is as follows:

- 0–4 mm: No pain
- 5–44 mm: Mild pain
- 45–74 mm: Moderate pain
- 75–100 mm: Severe pain

Advantages of the VAS include brevity, ease for patients and administrators, and comprehensive psychometric assessment.<sup>[22,39,47,53]</sup> In spine surgery, the VAS is commonly used to separately assess back pain (VAS-BP) and leg pain (VAS-LP). The highly subjective nature of pain limits comparison across a range of patients, however. In addition, it is highly dependent on the patient's short-term experience with pain, rather than a long-term average. Interestingly, slightly lower scores have been reported for horizontally oriented VAS compared to vertical ones.<sup>[39]</sup>

MCID: When studying emergency department patients with acute pain from primarily abdominal, extremity, and back etiologies, Lee *et al.* found an MCID of 3 mm.<sup>[49]</sup> Similarly, in patients undergoing surgical decompression and fusion for recurrent lumbar stenosis, Parker *et al.* found MCIDs of 2.2 and 5 for the VAS-BP and VAS-LP, respectively.<sup>[63]</sup>

### Numerical rating scale for pain

The numerical rating scale (NRS) is a one-dimensional scale for rating pain across multiple diseases and is essentially a variation of the VAS.<sup>[15,23,33]</sup> The most common format is a horizontal line with anchors at 0 and 10 (0 = no pain, 10 = worst pain imaginable), in which patients pick a discrete number on the scale (e.g., 3 or 4, not 3.5).<sup>[22,40]</sup> Like the VAS, most providers ask patients to rate their pain over the previous 24 hours.

Advantages of the NRS include brevity and published psychometrics.<sup>[22,29,40,41]</sup> In addition, one study found higher rates of data completion using the NRS compared to the VAS.<sup>[23]</sup> The primary criticism of the NRS, like the VAS, is that one-dimensional instruments fail to capture the true experience of pain and can mislead providers regarding patient outcomes.<sup>[38]</sup>

MCID: When studying the NRS using a 0–10 scale in clinical trials for diabetic neuropathy, postherpetic neuralgia, chronic low BP, fibromyalgia, and osteoarthritis,

Farrar *et al.* found a change of 2 points to be clinically meaningful.<sup>[28]</sup> Similar results were found when using a 15-point scale for patients with low BP.<sup>[15]</sup>

### McGill Pain Questionnaire

The McGill Pain Questionnaire (MPQ) was introduced in 1975, with a shorter version (SF-MPQ) developed in 1987.<sup>[57,58]</sup> The SF-MPQ measures pain experience using three components: A list of 15 pain descriptors, a “present pain intensity (PPI)” index, and a VAS. Each of the pain descriptors is assigned to progressive levels of pain (0 = none, 1 = mild, 2 = moderate, 3 = severe). In 2009, the SF-MPQ was modified (SF-MPQ-2) to include an additional seven neuropathic-specific markers and more responsive scoring methodology.<sup>[24]</sup> The SF-MPQ-2 also changed the answer range for each descriptor to 0–10 (0 = none, 10 = worst possible).

The SF-MPQ is scored from 0 to 45 for the pain descriptors, 0–5 for the PPI index, and on a continuum for the VAS. These values can be used as individual measures for each component or summed for total pain score. The scoring is slightly different for the SF-MPQ-2 because patients select a value from 0 to 10 for each descriptor, resulting in a pain descriptor range of 0–220.<sup>[24]</sup>

Advantages of the SF-MPQ are its comprehensiveness, ease of completion, and psychometric assessment.<sup>[13,32,50,79]</sup> However, Grafton *et al.* found that supervision and guidance are still needed for new users to adequately complete the SF-MPQ.<sup>[32]</sup>

MCID: Strand *et al.* found an MCID of >5 points on the 0–45 scale for patients with rheumatoid and musculoskeletal pain.<sup>[79]</sup> In the osteoarthritis population, Grafton *et al.* found an MCID of 5.2 for the total score, 4.5 for the sensory descriptors, 2.8 for the affective descriptors, 1.4 for the PPI, and 1.4 for the VAS.<sup>[32]</sup>

## LUMBAR SPINE PATHOLOGY

### Oswestry disability index

The Oswestry disability index (ODI) was first described in 1980 and measures functional disability from low BP.<sup>[26]</sup> It has become the most widely used PROM for BP and is most sensitive in patients with severe pain. It consists of 10 items related to everyday activities using a series of six short descriptions of progressively worse disability.

The ODI is scored from 0 to 100 (0 = no pain/disability, 100 = immobilizing pain/disability). Scoring is completed by summing the answers for each question on a 0–5 scale, dividing by the total possible score, and then multiplying by 100. Up to two questions may be omitted with preserved validity. The general interpretation of scores is as follows:

- 0–20: Minimal disability
- 21–40: Moderate disability

- 41–60: Severe disability
- 61–80: Crippled
- 81–100: Bedridden or exaggerating symptoms

The advantages of the ODI are simplicity, brevity, widespread use, psychometric validation, and sensitivity for higher levels of disability.<sup>[27]</sup> One disadvantage is that the ODI is relatively insensitive in patients with mild pain/disability compared to other instruments. In addition, the ODI is difficult to administer over the phone due to the length of the answer choices.

MCID: Using outcomes from the Lumbar Spine Study Group, Copay *et al.* determined an MCID of 12.8 for the ODI.<sup>[18,56]</sup> The Food and Drug Administration considers an MCID of 15 for patients undergoing spinal fusion.<sup>[27]</sup>

### Roland–Morris Disability Questionnaire

The Roland–Morris Disability Questionnaire (RMDQ) was developed in 1983, and unlike the ODI, is more sensitive in patients with mild/moderate BP.<sup>[69,70]</sup> It consists of 24 yes/no questions that begin with “because of my back pain.” Answers of “yes” indicate pain or disability. The RMDQ is scored from 0 to 24 by summing the number of “yes” responses to the 24 questions. Zero is considered no disability and 24 to be an immobilizing disability.

Advantages of the RMDQ are sensitivity to milder levels of low BP, simplicity, widespread use, and psychometric assessment.<sup>[9,42,48,51,65,69]</sup> Disadvantages include its length compared to the ODI, and that it does not address psychological or social disability. However, most patients complete the RMDQ in <5 minutes on average, despite having more than twice the questions of the ODI.<sup>[68-70]</sup>

MCID: After conducting a review of studies that used the RMDQ, Roland and Fairbank concluded that the MCID is between 2.5 and 5, depending on baseline disability.<sup>[68]</sup> Stratford *et al.* found MCIDs of 1–2 for patients with mild baseline disability, 7–8 for patients with moderate to severe baseline disability, and 5 for patients without a previous disability selection.<sup>[80]</sup>

### Quebec Back Pain Disability Scale

The Quebec Back Pain Disability Scale (QBPDS) was introduced in 1996 to quantify functional disability from BP.<sup>[45]</sup> The survey consists of 20 questions across six categories: Bed and rest, sitting and standing, ambulation, movement, bending, and handling of large items. Each question is rated on a 6-item Likert scale ranging from “not difficult at all” (0) to “unable to do” (5). The total is summed to produce a score from 0 to 100, where 0 is no disability and 100 is complete disability.

Advantages of the QBPDS are brevity, sensitivity to pelvic girdle pain and disability, and psychometric validation.<sup>[45,46,72]</sup> The primary disadvantage is that it is not widely used in the spinal surgery literature, and there are concerns that it is not as reliable as the ODI.<sup>[30]</sup>

MCID: In patients receiving physical therapy for BP, Davidson and Keating found an MCID of 19, while Fritz and Irrgang found an MCID of 15 in similar patients undergoing physical therapy.<sup>[20,30]</sup>

### Japanese Orthopedic Association Back Pain Evaluation Questionnaire

The Japanese Orthopedic Association Back Pain Evaluation Questionnaire (JOABPEQ) was first described in 2007 to measure functional disability due to low BP.<sup>[31]</sup> It consists of 25 questions across five categories: Low BP, lumbar function, walking ability, social life function, and mental health. Answers choices are either yes/no or on a Likert scale. Scoring is completed using a unique weighting methodology for each answer and sub-category, the algorithm for which is available on the JOA website. The total score ranges from 0 to 20 for each of the five categories, with lower scores indicating more severe disability. The survey is not designed to produce an aggregate score of the five categories.

Advantages of the JOABPEQ include its breadth of measurement for each domain, weighting of questions, and psychometric validation.<sup>[4,31]</sup> However, these same benefits make the questionnaire much longer for patients to complete and more difficult to score and interpret. In addition, it is not widely used in the spine surgery literature.

MCID: Not determined.

## CERVICAL SPINE PATHOLOGY

### Neck disability index

The Neck disability index (NDI) was developed in 1991 as a modification of the ODI and serves to quantify disability due to neck pain.<sup>[84]</sup> It consists of 10 items related to pain, personal care, lifting, reading, headaches, concentration, work, driving, sleeping, and recreation. Each question is scored from 0 (no disability) to 5 (complete disability) and is then totaled (0–50). The score may be reported as either a raw score or percentage. If more than two questions are unanswered, the survey may no longer be valid. The general interpretation of scores is as follows:

- 0–8%: No disability
- 10–28%: Mild disability
- 30–48%: Moderate disability
- 50–64%: Severe disability
- 70–100%: Complete disability

Advantages of the NDI include simplicity, brevity, widespread use, and durability. It does not appear to suffer from floor or ceiling effects when used as intended, although floor and ceiling effects were observed when attempts were made to map the NDI to the SF-6D.<sup>[66]</sup>

One disadvantage is that there is no incorporation of medication use in the scoring.

MCID: Auffinger *et al.* found an MCID of 2.41 points following surgical treatment of degenerative cervical spine disease.<sup>[2]</sup> Carreon *et al.* found an MCID of 7.5 points following cervical fusion surgery, Cleland *et al.* found an MCID of 19% (9.5 points) in nonoperative neck pain patients undergoing physical therapy, and Auffinger *et al.* found an MCID of 13.39 points in surgically treated cervical myelopathy patients.<sup>[3,11,17]</sup> The wide range of reported MCIDs, differences in study methodology, and range of sample sizes make MCID cut-offs difficult for this instrument.

### Cervical Spine Outcomes Questionnaire

The Cervical Spine Outcomes Questionnaire (CSOQ) was developed in 2002 as a disease-specific measure of disability from cervical spine pathology.<sup>[7]</sup> 35 items are presented over six domains on neck pain, shoulder/arm pain, functional disability, psychological distress, physical symptoms, and healthcare utilization. The individual domains are scaled from 0 to 100 and intended for individual comparison rather than a cumulative score.

Advantages include published psychometrics, differentiation between neck and extremity pain, and the incorporation of pain medication use in the score.<sup>[76]</sup> Disadvantages include ceiling effects in physical symptoms,<sup>[76]</sup> and more importantly, it is uncommonly used and difficult to obtain (both the survey and scoring algorithm). In addition, it requires more time and effort by patients to complete compared to the NDI. The original survey was developed using surgical patients; thus, it may not be generalizable to nonsurgical patients, and details of the scoring and weighting processes are unclear from the original manuscript.

MCID: Skolasky *et al.* assessed the CSOQ on anterior cervical discectomy and fusion patients and found MCIDs of 0.13–0.24 for the various domains.<sup>[75]</sup> It should be noted that scores for the CSOQ in that study appear to be on a 0–1 scale, rather than the 0–100 scale originally described.

### Modified Japanese Orthopedic Association Scale

The Japanese Orthopedic Association Scale (JOA) was developed in 1985 to quantify the severity of cervical myelopathy.<sup>[36]</sup> It consists of 6 items in three domains: Motor function (upper extremity, lower extremity), sensory function (upper extremity, trunk, lower extremity), and bladder function. Because the original survey is geared toward a Japanese audience (e.g., upper extremity dysfunction is evaluated by the ability to use chopsticks), it was modified in 1991 by Benzel *et al.* to consolidate the survey to four questions and contextually translate it for Western audiences.<sup>[8]</sup> The second

modification was introduced by Chiles *et al.* that reverted to the original 6 items while maintaining the Western contextual translation,<sup>[16]</sup> although Benzel's version remains the more commonly used.

Scoring depends on which modified JOA (mJOA) version is used; both versions are additive, but Benzel's version is scored out of 18 total points, while the Chiles version is scored out of 17 points. Lower scores indicate more severe myelopathy. Although there are no formal strata, the literature suggests scores >12 indicate mild myelopathy.<sup>[52]</sup>

Advantages of the mJOA are simplicity, brevity, widespread use, and durability. The primary disadvantage is that it has never been psychometrically validated despite widespread acceptance and recommendation by professional societies.

MCID: Not determined.

### Myelopathy disability index

The myelopathy disability index was developed in 1996 as a modified version of the Stanford Health Assessment Questionnaire.<sup>[12]</sup> It was designed to evaluate the degree of cervical myelopathy in rheumatoid arthritis patients undergoing surgical intervention for cervical myelopathy. The survey has 10 items regarding standing, walking, grip strength, eating, and hygiene, each with four choices (0–3). The results are additive (maximum score of 30), with the final score generally expressed as a percentage. Higher scores indicate more severe myelopathy.

Advantages include its brevity and psychometric validation.<sup>[12]</sup> However, it is not commonly used compared to the mJOA scale and seems to have fallen out of favor in recent years.

MCID: Not determined.

Finally, the Nurick scale is a common instrument for assessing cervical myelopathy. While some may interpret the Nurick scale as a PROM, it is a physician-assigned instrument, as lower grades on the scale are based on signs of root or cord involvement rather than patient-endorsed symptoms.

## COSTS

A brief overview on costs is relevant given the intrinsic relationship between PROMs and cost-effectiveness research. There are three principal types of costs: Direct costs (e.g., physician time and labor, cost of operating room, length of stay, etc.), indirect costs (e.g., lost productivity from missed work days), and intangible costs (e.g., pain and suffering). For the purposes of cost-utility analysis in spine surgery, most researchers attempt to use a "societal perspective," which is the sum of all direct and indirect costs.

Direct costs include inpatient hospital costs, outpatient expenses, and physician professional fees. There are numerous methods with which to estimate costs, most of which rely on hospital charge data, hospital reimbursement data, or average payer reimbursement rates.<sup>[1]</sup> Unfortunately, methods of estimating costs are often imprecise and may yield highly discrepant data, particularly between countries with varying health care systems. For example, Whitmore *et al.* compared estimated direct hospital costs for the surgical treatment of cervical myelopathy in US patients using two methods: (1) Based on hospital charge data, and (2) based on average Medicare reimbursements.<sup>[90]</sup> For dorsal surgery, the estimated cost of the index hospitalization using the first method was nearly \$11,000 higher than the second method. In addition, dorsal procedures appeared to be more expensive than ventral procedures in the first method, while the second method suggested the opposite. The authors concluded that the choice of cost methodology substantially affects the final results of cost-utility studies.

Additional costs that are often overlooked include re-admission/re-operation and outpatient expenses. Patients may be re-admitted for a variety of reasons, such as a deep vein thrombosis requiring treatment, or may require early or delayed re-operation for reasons related to the index procedure, such as postoperative hematoma or hardware failure. These complications can multiply the aggregate costs of care. Therefore, data related to re-operation and re-admission should be tracked prospectively. Outpatient expenses, such as physical therapy, medication use, and radiographic screening, may become substantial yet are frequently omitted due to the difficulties in accurately tracking outpatient resource utilization. With regard to indirect costs, time off from work is the central variable and should ideally be tracked prospectively to minimize recall bias. Indirect costs are estimated by multiplying a patient's wages by time off from work, although national wage indices are often used as a proxy if actual wage data are unavailable.

Once both costs and outcomes (QALYs) are measured, the results may be presented as the cost per QALY gained for a given procedure. When comparing two procedures, the incremental cost-effectiveness ratio (ICER) can be obtained using the formula:  $(\text{Cost of treatment A} - \text{Cost of Treatment B}) / (\text{QALY of treatment A} - \text{QALY of treatment B})$ . Lower ratios imply better cost-effectiveness. The ICER needs to be interpreted in the context of society's willingness to pay. Previous studies have suggested a seemingly arbitrary cost-effectiveness ratio <\$50,000/QALY as acceptable in many developed countries, although the figure is highly contentious, with some authors suggesting significantly less, others significantly more, and some concluding a mismatch

between willingness to pay theory and usefulness in actual practice.<sup>[34,44,61]</sup>

## DISCUSSION

Given the focus on patient-centered medical care and increased attention to the value of healthcare, PROMs will be increasingly used in clinical research moving forward. These instruments attempt to quantify and provide uniformity in evaluating characteristics that clinicians have traditionally assessed subjectively. While they are imperfect, PROMs are the building blocks for our ability to improve care and assess the impact of surgical interventions from a patient's perspective.

The choice of instrument selection depends on several factors, including psychometric evaluation, practicality to the patient, administrative burden, use in published literature, and professional consensus, among others. Most spinal diseases cause a combination of pain and disability, which in turn affects QoL. Thus, at least one instrument from each major category should be employed in the evaluation of spine patients. In nonspinal diseases, however, the core pathology may be completely asymptomatic yet still require treatment (e.g., unruptured intracranial aneurysms). In such patient population, determining the appropriate outcome measures can be more challenging. Instruments should be administered preoperatively and at scheduled time points postoperatively. Preoperative baseline values and serial postoperative scores are particularly important for generic HRQoL instruments because determining QALYs requires both time points and is essential for conducting comparative- and cost-effectiveness research.

Patient and administrative burden is particularly important to consider with the rise of prospective registries, in which patients are followed for long periods of time postoperatively. Longer surveys have been shown to have lower completion rates, which detract from the overall quality of data.<sup>[6,35]</sup> Perhaps for that reason, the EQ-5D has been adopted as the generic HRQoL outcome for both the National Neurosurgery Quality and Outcomes Database and the UK NICE quality improvement initiative, and will likely become even more commonly encountered in the future.<sup>[14,54]</sup> Finally, it should be re-iterated that SF-6D and EQ-5D surveys often derive very different utility values for the same condition, thus when comparing multiple studies for a given disease or surgery, it is important to keep in mind which instrument was employed.<sup>[89]</sup>

For disease-specific instruments, the MCID, if determined, should be considered when evaluating outcomes. Although there may be statistically significant changes detected following an intervention, it may not be clinically meaningful. Most studies on MCID have used very

specific patient population with small sample sizes; thus, the MCIDs published to date need to be interpreted with caution. In addition, it should be noted that some disease-specific instruments that are widely accepted and commonly used may have not have undergone rigorous psychometric assessment, an opportunity for future research. Finally, while they do quantify treatment effect, QALYs are not derived from disease-specific instruments, and, therefore, cannot be relied upon for cost-utility studies.

Moving forward, increased uniformity among PROMs in published studies would improve the comparability within health services research. Furthermore, there is a continued need for disease-specific instruments for common subsets of spine surgery patients, most notably radiculopathy. Most cases of radiculopathy do not require surgery, and given the highly subjective nature of pain, it is difficult to track the true functional debilitation from radiculopathy aside from proxy use of the ODI/NDI or HRQoL instruments. The US National Institutes of Health recently allocated resources for the development of new, validated PROMs through a program entitled PROM Information Systems (PROMIS), which uses item response theory to develop shorter, psychometrically validated instruments. One recent study has already found the PROMIS physical functioning survey to be appropriate in patients with degenerative spine conditions.<sup>[37]</sup> However, it will be interesting to see whether this program consolidates the plethora of instruments currently in use or only expands it.

## CONCLUSION

PROMs measure important aspects of disease burden from a patient's perspective. While traditional outcomes measures in spine surgery have relied on clinical data, radiographic studies, and physician-assigned scales, there has been an increasing trend for the use of PROMs in routine clinical practice, which will likely become standard of care in the future. The broad categories of PROMs include generic HRQoL, pain, and disease-specific disability, all of which should be measured preoperatively and at regular intervals postoperatively to obtain a comprehensive assessment of the patient's condition. The foremost goal is to select appropriate instruments given the patient's underlying condition, but one should also keep in mind the sustainability of measuring outcomes from patient burden and administrative standpoints. In addition, the accurate estimation of costs is a challenging but necessary task as increased attention is paid to the relative value of surgical interventions.

### Financial support and sponsorship

Nil.

### Conflicts of interest

There are no conflicts of interest.



## REFERENCES

- Alvin MD, Miller JA, Lubelski D, Rosenbaum BP, Abdullah KG, Whitmore RG, et al. Variations in cost calculations in spine surgery cost-effectiveness research. *Neurosurg Focus* 2014;36:E1.
- Auffinger B, Lam S, Shen J, Thaci B, Roitberg BZ. Usefulness of minimum clinically important difference for assessing patients with subaxial degenerative cervical spine disease: Statistical versus substantial clinical benefit. *Acta Neurochir (Wien)* 2013;155:2345-54.
- 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:e67408.
- Azimi P, Shahzadi S, Montazeri A. The Japanese Orthopedic Association Back Pain Evaluation Questionnaire (JOABPEQ) for low back disorders: A validation study from Iran. *J Orthop Sci* 2012;17:521-5.
- Baron R, Elashaal A, Germon T, Hobart J. Measuring outcomes in cervical spine surgery: Think twice before using the SF-36. *Spine (Phila Pa 1976)* 2006;31:2575-84.
- Barton GR, Sach TH, Avery AJ, Jenkinson C, Doherty M, Whyne DK, et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged >or=45 years. *Health Econ* 2008;17:815-32.
- BenDebba M, Heller J, Ducker TB, Eisinger JM. Cervical spine outcomes questionnaire: Its development and psychometric properties. *Spine (Phila Pa 1976)* 2002;27:2116-23.
- Benzel EC, Lancon J, Kesterson L, Hadden T. Cervical laminectomy and dentate ligament section for cervical spondylotic myelopathy. *J Spinal Disord* 1991;4:286-95.
- Beurskens AJ, de Vet HC, K  ke AJ. Responsiveness of functional status in low back pain: A comparison of different instruments. *Pain* 1996;65:71-6.
- Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Qual Life Res* 1993;2:169-80.
- 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;10:469-74.
- Casey AT, Bland JM, Crookard HA. Development of a functional scoring system for rheumatoid arthritis patients with cervical myelopathy. *Ann Rheum Dis* 1996;55:901-6.
- Chaffee A, Yakuboff M, Tanabe T. Responsiveness of the VAS and McGill pain questionnaire in measuring changes in musculoskeletal pain. *J Sport Rehabil* 2011;20:250-5.
- Chard J, Kuczawski M, Black N, van der Meulen J, Committee POAS. Outcomes of elective surgery undertaken in independent sector treatment centres and NHS providers in England: Audit of patient outcomes in surgery. *Br Med J* 2011;343:d6404.
- Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. *Spine (Phila Pa 1976)* 2005;30:1331-4.
- Chiles BW 3rd, Leonard MA, Choudhri HF, Cooper PR. Cervical spondylotic myelopathy: Patterns of neurological deficit and recovery after anterior cervical decompression. *Neurosurgery* 1999;44:762-9.
- Cleland JA, Childs JD, Whitman JM. Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. *Arch Phys Med Rehabil* 2008;89:69-74.
- Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: A choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. *Spine J* 2008;8:968-74.
- Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: A review of concepts and methods. *Spine J* 2007;7:541-6.
- Davidson M, Keating JL. A comparison of five low back disability questionnaires: Reliability and responsiveness. *Phys Ther* 2002;82:8-24.
- Devlin NJ, Parkin D, Browne J. Patient-reported outcome measures in the NHS: New methods for analysing and reporting EQ-5D data. *Health Econ* 2010;19:886-905.
- Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. *Ann Rheum Dis* 1978;37:378-81.
- Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005;113:9-19.
- Dworkin RH, Turk DC, Revicki DA, Harding G, Coyne KS, Peirce-Sandner S, et al. Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2). *Pain* 2009;144:35-42.
- EuroQol Group. EuroQol - A new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
- Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271-3.
- Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)* 2000;25:2940-52.
- Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001;94:149-58.
- Ferraz MB, Quaresma MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. *J Rheumatol* 1990;17:1022-4.
- Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther* 2001;81:776-88.
- Fukui M, Chiba K, Kawakami M, Kikuchi S, Konno S, Miyamoto M, et al. JOA Back Pain Evaluation Questionnaire (JOABPEQ)/JOA Cervical Myelopathy Evaluation Questionnaire (JOACMEQ). The report on the development of revised versions. April 16, 2007. The Subcommittee of the Clinical Outcome Committee of the Japanese Orthopaedic Association on Low Back Pain and Cervical Myelopathy Evaluation. *J Orthop Sci* 2009;14:348-65.
- Grafton KV, Foster NE, Wright CC. Test-retest reliability of the Short-Form McGill Pain Questionnaire: Assessment of intraclass correlation coefficients and limits of agreement in patients with osteoarthritis. *Clin J Pain* 2005;21:73-82.
- Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short-Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res (Hoboken)* 2011;63 Suppl 11:S240-52.
- Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality-adjusted life year: In search of a standard. *Med Decis Making* 2000;20:332-42.
- Holland R, Smith RD, Harvey I, Swift L, Lenaghan E. Assessing quality of life in the elderly: A direct comparison of the EQ-5D and AQL. *Health Econ* 2004;13:793-805.
- Hukuda S, Mochizuki T, Ogata M, Shichikawa K, Shimomura Y. Operations for cervical spondylotic myelopathy. A comparison of the results of anterior and posterior procedures. *J Bone Joint Surg Br* 1985;67:609-15.
- Hung M, Hon SD, Franklin JD, Kendall RV, Lawrence BD, Neese A, et al. Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. *Spine (Phila Pa 1976)* 2014;39:158-63.
- Hush JM, Refshauge KM, Sullivan G, De Souza L, McAuley JH. Do numerical rating scales and the Roland-Morris Disability Questionnaire capture changes that are meaningful to patients with persistent back pain? *Clin Rehabil* 2010;24:648-57.
- Huskisson EC. Measurement of pain. *Lancet* 1974;2:1127-31.
- Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: A comparison of six methods. *Pain* 1986;27:117-26.
- Jensen MP, Karoly P, O'Riordan EF, Bland F Jr, Burns RS. The subjective experience of acute pain. An assessment of the utility of 10 indices. *Clin J Pain* 1989;5:153-9.
- Jensen MP, Strom SE, Turner JA, Romano JM. Validity of the Sickness Impact Profile Roland scale as a measure of dysfunction in chronic pain patients. *Pain* 1992;50:157-62.
- Kahn TL, Soheili A, Schwarzkopf R. Outcomes of total knee arthroplasty in relation to preoperative patient-reported and radiographic measures: Data from the osteoarthritis initiative. *Geriatr Orthop Surg Rehabil* 2013;4:117-26.
- King JT Jr, Tsevat J, Lave JR, Roberts MS. Willingness to pay for a quality-adjusted life year: Implications for societal health care resource allocation. *Med Decis Making* 2005;25:667-77.
- Kopec JA, Esdaile JM, Abrahamowicz M, Abenham L, Wood-Dauphinee S,

- Lamping DL, et al. The Quebec Back Pain Disability Scale. Measurement properties. *Spine (Phila Pa 1976)* 1995;20:341-52.
46. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al. The Quebec Back Pain Disability Scale: Conceptualization and development. *J Clin Epidemiol* 1996;49:151-61.
  47. Langley GB, Sheppard H. The visual analogue scale: Its use in pain measurement. *Rheumatol Int* 1985;5:145-8.
  48. Leclaire R, Blier F, Fortin L, Proulx R. A cross-sectional study comparing the Oswestry and Roland-Morris Functional Disability scales in two populations of patients with low back pain of different levels of severity. *Spine (Phila Pa 1976)* 1997;22:68-71.
  49. Lee JS, Hobden E, Stiell IG, Wells GA. Clinically important change in the visual analog scale after adequate pain control. *Acad Emerg Med* 2003;10:1128-30.
  50. Lovejoy TI, Turk DC, Morasco BJ. Evaluation of the psychometric properties of the revised short-form McGill Pain Questionnaire. *J Pain* 2012;13:1250-7.
  51. Macedo LG, Maher CG, Latimer J, Hancock MJ, Machado LA, McAuley JH. Responsiveness of the 24-, 18- and 11-item versions of the Roland Morris Disability Questionnaire. *Eur Spine J* 2011;20:458-63.
  52. Matz PG, Anderson PA, Holly LT, Groff MW, Heary RF, Kaiser MG, et al. The natural history of cervical spondylotic myelopathy. *J Neurosurg Spine* 2009;11:104-11.
  53. McCormack HM, Horne DJ, Sheather S. Clinical applications of visual analogue scales: A critical review. *Psychol Med* 1988;18:1007-19.
  54. McGirt MJ, Speroff T, Dittus RS, Harrell FE Jr, Asher AL. The National Neurosurgery Quality and Outcomes Database (N2QOD): General overview and pilot-year project description. *Neurosurg Focus* 2013;34:E6.
  55. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Qual Life Res* 1995;4:293-307.
  56. Meade TW, Dyer S, Browne W, Townsend J, Frank AO. Low back pain of mechanical origin: Randomized comparison of chiropractic and hospital outpatient treatment. *J Orthop Sports Phys Ther* 1991;13:278-87.
  57. Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods. *Pain* 1975;1:277-99.
  58. Melzack R. The short-form McGill Pain Questionnaire. *Pain* 1987;30:191-7.
  59. Methodology Committee of the Patient-Centered Outcomes Research I. Methodological standards and patient-centeredness in comparative effectiveness research: The PCORI perspective. *J Am Med Assoc* 2012;307:1636-40.
  60. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-45.
  61. Olsen JA, Smith RD. Theory versus practice: A review of 'willingness-to-pay' in health and health care. *Health Econ* 2001;10:39-52.
  62. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: Towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 2008;33:90-4.
  63. Parker SL, Adogwa O, Paul AR, Anderson WN, Aaronson O, Cheng JS, et al. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. *J Neurosurg Spine* 2011;14:598-604.
  64. Parker SL, Godil SS, Shau DN, Mendenhall SK, McGirt MJ. Assessment of the minimum clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion: Clinical article. *J Neurosurg Spine* 2013;18:154-60.
  65. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. *Spine (Phila Pa 1976)* 1995;20:1899-908.
  66. Richardson SS, Berven S. The development of a model for translation of the Neck Disability Index to utility scores for cost-utility analysis in cervical disorders. *Spine J* 2012;22:55-62.
  67. Roguski M, Benzel EC, Curran JN, Magge SN, Bisson EF, Krishnaney AA, et al. Postoperative cervical sagittal imbalance negatively affects outcomes after surgery for cervical spondylotic myelopathy. *Spine (Phila Pa 1976)* 2014;39:2070-7.
  68. Roland M, Fairbank J. The Roland-Morris disability questionnaire and the Oswestry disability questionnaire. *Spine (Phila Pa 1976)* 2000;25:3115-24.
  69. Roland M, Morris R. A study of the natural history of back pain. Part I: Development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976)* 1983;8:141-4.
  70. Roland M, Morris R. A study of the natural history of low-back pain. Part II: Development of guidelines for trials of treatment in primary care. *Spine (Phila Pa 1976)* 1983;8:145-50.
  71. Sach TH, Barton GR, Jenkinson C, Doherty M, Avery AJ, Muir KR. Comparing cost-utility estimates: Does the choice of EQ-5D or SF-6D matter? *Med Care* 2009;47:889-94.
  72. Schoppink LE, van Tulder MW, Koes BW, Beurskens SA, de Bie RA. Reliability and validity of the Dutch adaptation of the Quebec Back Pain Disability Scale. *Phys Ther* 1996;76:268-75.
  73. Services USDoHaH. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. *Health Qual Life Outcomes* 2006;4:79.
  74. Shaw FE, Asomugha CN, Conway PH, Rein AS. The Patient Protection and Affordable Care Act: Opportunities for prevention and public health. *Lancet* 2014;384:75-82.
  75. Skolasky RL, Albert TJ, Maggard AM, Riley LH 3<sup>rd</sup>. Minimum clinically important differences in the Cervical Spine Outcomes Questionnaire: Results from a national multicenter study of patients treated with anterior cervical decompression and arthrodesis. *J Bone Joint Surg Am* 2011;93:1294-300.
  76. Skolasky RL, Riley LH 3<sup>rd</sup>, Albert TJ. Psychometric properties of the Cervical Spine Outcomes Questionnaire and its relationship to standard assessment tools used in spine research. *Spine J* 2007;7:174-9.
  77. Søgaard R, Christensen FB, Videbaek TS, Bünger C, Christiansen T. Interchangeability of the EQ-5D and the SF-6D in long-lasting low back pain. *Value Health* 2009;12:606-12.
  78. Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, et al. The impact of patient-centered care on outcomes. *J Fam Pract* 2000;49:796-804.
  79. Strand LI, Ljunggren AE, Bogen B, Ask T, Johnsen TB. The Short-Form McGill Pain Questionnaire as an outcome measure: Test-retest reliability and responsiveness to change. *Eur J Pain* 2008;12:917-25.
  80. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part I. *Phys Ther* 1998;78:1186-96.
  81. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
  82. Tosteson AN, Skinner JS, Tosteson TD, Lurie JD, Andersson GB, Berven S, et al. The cost effectiveness of surgical versus nonoperative treatment for lumbar disc herniation over two years: Evidence from the Spine Patient Outcomes Research Trial (SPORT). *Spine (Phila Pa 1976)* 2008;33:2108-15.
  83. van Stel HF, Buskens E. Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease. *Health Qual Life Outcomes* 2006;4:20.
  84. Vernon H, Mior S. The Neck Disability Index: A study of reliability and validity. *J Manipulative Physiol Ther* 1991;14:409-15.
  85. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;14:1523-32.
  86. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
  87. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, et al. Surgical vs nonoperative treatment for lumbar disk herniation: The Spine Patient Outcomes Research Trial (SPORT): A randomized trial. *JAMA* 2006;296:2441-50.
  88. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 1996;276:1253-8.
  89. Whitehurst DG, Bryan S, Lewis M. Systematic review and empirical comparison of contemporaneous EQ-5D and SF-6D group mean scores. *Med Decis Making* 2011;31:E34-44.
  90. Whitmore RG, Schwartz JS, Simmons S, Stein SC, Ghogawala Z. Performing a cost analysis in spine outcomes research: Comparing ventral and dorsal approaches for cervical spondylotic myelopathy. *Neurosurgery* 2012;70:860-7.