J Neurosurg Spine 21:7–13, 2014

Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 2: Assessment of functional outcome following lumbar fusion

Zoher Ghogawala, M.D.,[1] Daniel K. Resnick, M.D.,[2] William C. Watters III, M.D.,[3] Praveen V. Mummaneni, M.D.,[4] Andrew T. Dailey, M.D.,[5] Tanvir F. Choudhri, M.D.,[6] Jason C. Eck, D.O., M.S.,[7] Alok Sharan, M.D.,[8] Michael W. Groff, M.D.,[9] Jeffrey C. Wang, M.D.,[10] Sanjay S. Dhall, M.D.,[4] and Michael G. Kaiser, M.D.[11]

[1]Alan and Jacqueline Stuart Spine Research Center, Department of Neurosurgery, Lahey Clinic, Burlington, and Tufts University School of Medicine, Boston, Massachusetts; [2]Department of Neurosurgery, University of Wisconsin, Madison, Wisconsin; [3]Bone and Joint Clinic of Houston, Texas; [4]Department of Neurological Surgery, University of California, San Francisco, California; [5]Department of Neurosurgery, University of Utah, Salt Lake City, Utah; [6]Department of Neurosurgery, Icahn School of Medicine at Mount Sinai, New York, New York; [7]Center for Sports Medicine and Orthopaedics, Chattanooga, Tennessee; [8]Department of Orthopaedic Surgery, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York; [9]Department of Neurosurgery, Brigham and Women's Hospital, Boston, Massachusetts; [10]Department of Orthopaedic Surgery, Keck School of Medicine, University of Southern California, Los Angeles, California; and [11]Department of Neurosurgery, Columbia University, New York, New York

Assessment of functional patient-reported outcome following lumbar spinal fusion continues to be essential for comparing the effectiveness of different treatments for patients presenting with degenerative disease of the lumbar spine. When assessing functional outcome in patients being treated with lumbar spinal fusion, a reliable, valid, and responsive outcomes instrument such as the Oswestry Disability Index should be used. The SF-36 and the SF-12 have emerged as dominant measures of general health-related quality of life. Research has established the minimum clinically important difference for major functional outcomes measures, and this should be considered when assessing clinical outcome. The results of recent studies suggest that a patient's pretreatment psychological state is a major independent variable that affects the ability to detect change in functional outcome.
(http://thejns.org/doi/abs/10.3171/2014.4.SPINE14258)

Key Words • **fusion** • **lumbar spine** • **treatment outcomes** • **practice guidelines**

Recommendations

There is no evidence that conflicts with the previous recommendations published in the original version of the "Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine."

Grade B

It is recommended that when assessing functional outcome in patients treated for low-back pain due to degenerative disease, a reliable, valid, and responsive outcomes instrument, such as the disease-specific Oswestry Disability Index (ODI), be used (Level II evidence).

It is recommended that when assessing general health-related quality of life (HR-QOL) in patients treated for low-back pain due to degenerative disease that a reliable, valid, and responsive outcomes instrument, such as the 36-Item Short Form Health Survey (SF-36), be used (Level II evidence).

Abbreviations used in this paper: BIS = Balanced Inventory for Spinal Disorders; DRI = Disability Rating Index; HR-QOL = health-related quality of life; ICC = intraclass correlation coefficient; LSOQ = Lumbar Spine Outcomes Questionnaire; MCID = minimum clinically important difference; MCS = mental component summary; ODI = Oswestry Disability Index; PCS = physical component summary; RMDQ = Roland-Morris Disability Questionnaire; ROC = receiver operating characteristic; SF-12 = 12-Item Short Form Health Survey; SF-36 = 36-Item Short Form Health Survey; SIP = Sickness Impact Profile; SRS-22 = 22-Item Scoliosis Research Society; VAS = visual analog scale.

It is recommended that the minimum clinically important difference (MCID) be considered when assessing clinical outcome (Level II evidence).

Rationale

The assessment of functional outcome for patients who undergo lumbar fusion surgery continues to be an area of intense clinical interest. The Institute of Medicine has identified low-back pain treatment options as one of the highest priorities for new comparative effectiveness research.[11] In an effort to improve the reporting of outcomes following lumbar fusion, an emphasis has been placed on the implementation of valid, reliable, and objective outcome measures. The majority of these instruments are patient self-assessment questionnaires that report quality of life. They can be divided into 2 groups: those that seek to measure disease-specific outcomes, such as the ODI, and general health surveys, such as the SF-36.

The original Lumbar Fusion Guidelines recommended the utilization of reliable, valid, and responsive instruments to assess clinical outcome following treatment for low-back pain; however, there was insufficient evidence to standardize the utilization of one instrument over another, and multiple options were suggested. Patient satisfaction scales, however, were discouraged unless no alternative was available. Since the publication of the first generation of guidelines, investigators have continued to evaluate the utility of these instruments in the assessment of patients treated for low-back pain. We have assessed functional outcome measures by evaluating the evidence from a diagnostic perspective. That is, measurement of functional outcome would not be expected to improve outcome per se, but rather should allow investigators to "diagnose" any improvement in outcome following treatment.

Search Criteria

A computerized search of the National Library of Medicine database of the literature published between 2004 and 2011 was performed. The following subject headings and configurations yielded 1297 citations: ((("Lumbosacral Region"[MeSH] OR "Lumbar Vertebrae"[MeSH]) AND "Spinal Fusion"[MeSH]) OR "lumbar fusion"[All Fields] OR ("lumbar"[title] AND "fusion"[title])) AND ("Treatment Outcome"[MeSH] OR "Patient Satisfaction"[MeSH] OR "functional outcome"[All Fields] OR "functional outcomes"[All Fields] OR "outcome"[title] OR "outcomes"[title]). An additional search using "lumbar spine surgery," "outcomes," and "validation studies" yielded an additional 11 citations. The titles and abstracts of the 1308 articles were reviewed, and 28 clinical series focusing on adult patients who underwent lumbar fusion procedures were selected for analysis. Among the articles reviewed from this search, 10 have been included in the evidentiary table (see Table 1) along with 5 major articles (Level II evidence) from the original Lumbar Fusion Guidelines.[17] These 15 articles form the basis for these recommendations. Two studies focused on the reliability of new outcome measures. Four studies examined the reliability, validity, and responsiveness of a new lumbar spine outcomes measure. Four additional studies focused on the validity of established lumbar spine outcome measures, 1 study examined the responsiveness of a specific outcome measure, and 1 study calculated MCIDs for 4 major lumbar spine outcome measures. Three studies reported major predictors of functional outcome for lumbar spine patients. Among the 15 studies, 14 studies provided Level II and 1 study provided Level III medical evidence regarding functional outcome measures from a diagnostic perspective.

Scientific Foundation

Characteristics of a Functional Outcome Instrument

The criteria that determine whether a functional outcome instrument appropriately measures the response to treatment have not changed since the publication of the original guidelines in 2005.[17] The accuracy of an outcome instrument is dependent on 3 qualities—reliability, validity, and responsiveness.[6,7,13] Reliability is the measure of an instrument's consistency or reproducibility when reporting observations and is described by the following characteristics: interobserver reliability (the degree to which different observers obtain similar results when measuring the same phenomenon), intraobserver reliability (the extent to which the same observer obtains similar results on repeated observations of a fixed characteristic), test-retest reliability (consistency of an instrument between 2 separate time points, similar to intraobserver reliability, except that the characteristic, if clinical, may change with time), and internal consistency (used to describe the extent to which individual test domains correlate with the composite result).[12]

Reliability of an instrument is measured statistically in a variety of ways, depending on the nature of the recording of the observation: the κ statistic measures agreement between observers or observations beyond chance when the measure is in the form of categorical data, phi is used with dichotomous data, and intraclass correlation coefficient (ICC) is used with continuous data (and can be used with categorical data). In addition, the α statistic is used to measure internal consistency—the degree to which individual aspects (called "domains") of an outcome measure correlate with the composite result. A functional outcome measure is considered highly reliable if the κ value is greater than 0.8. A measure is thought to be moderately reliable if the κ value is between 0.6 and 0.8. A κ value of less than 0.6 suggests that the outcome measure is less reliable.[10] The internal reliability (α) is generally measured using the Cronbach α test to determine whether individual domains of a test correlate with the final composite result.[4]

The second criterion used to evaluate a functional outcome measure is validity, the ability to measure the disease-specific properties of interest. More recent literature compares novel functional outcome measures with previously validated instruments to assess validity.[15] Typically, the Pearson product-moment coefficient of correlation (r) is used to examine the congruency between one outcome measure and another, with $r > 0.80$ representing a strong correlation between measures.[14] Newer measures, such as the 22-Item Scoliosis Research Soci-

TABLE 1: Assessment of functional outcome: summary of evidence*

Authors & Year	Level of Evidence	Study Description	Comments
Fairbank et al., 1980	II	25 pts w/ acute LBP & reasonable prognosis were studied at weekly intervals for 3 wks w/ a functional disability survey. The ODI has 10 categories, each w/ 6 responses graded 0–5. A total of 50 points are possible. Test-retest reliability was κ >0.95 ($p < 0.001$) in 22 pts. Over the 3-wk interval, significant improvement was noted clinically & was detected using the ODI. A paired t-test revealed significant improvement on the ODI over 3 wks ($p < 0.05$).	The ODI is a reliable, valid, & responsive measure for detecting changes in LBP & its functional severity.
Roland & Morris, 1983	II	230 pts w/ acute LBP; 193 were studied at 0, 1, & 4 wks after the episode. Test-retest reliability was done in 20 of 230 cases. The construct validity was qualitatively assessed by comparing the functional questionnaire to the pain rating scale. External reliability was κ >0.90 & internal consistency was α >0.80. Construct validity demonstrated that the RMDQ was able to qualitatively detect pts w/ poorer outcomes from acute lumbago; however, no specific analysis was done.	The RMDQ is reliable for assessment of acute LBP.
Deyo, 1986	II	136 pts were examined in a clinic for a chief complaint of LBP. Evaluation was done using SIP & the modified RMDQ (shortened version of SIP) initially & 3 weeks later. Reliability for both scales was κ >0.80 in pts (n = 10) who had no change in pain. For pts who did not resume full activity (n = 47), the reliability was α >0.60. A strong correlation existed btwn the scales (r = 0.85) & btwn the physical dimension of the SIP & the modified RMDQ (r = 0.89). The modified RMDQ correlated less well w/ the psychosocial dimension of the SIP (r = 0.56).	The SIP & modified RMDQ (shorter) are reliable scales for the assessment of LBP that seem to follow the physical dimension of functional disability. The modified RMDQ is less well suited to follow the psychosocial dimension of functional disability.
Salén et al., 1994	II	1445 participants were divided into 3 groups: 1092 volunteer controls, 306 pts w/ axial skeletal pain, & 47 w/ joint pain. Patients were evaluated using the DRI & an FSQ. External reliability for the DRI was κ >0.80. There was a correlation to the FSQ. The DRI was responsive in detecting improvement after joint replacement.	The DRI is a reliable, valid, & responsive measure in pts w/ axial skeletal pain.
Luo et al., 2003	II	Study of 2520 pts w/ LBP; 506 pts assessed over 3–6 months using SF-12. The reliability coefficient for PCS-12 was 0.77 & for MCS-12 it was 0.80. Validity was verified by correlating results w/ back pain intensity & ODI. Responsiveness was verified.	The SF-12 is a reliable, valid, & responsive outcomes instrument for pts w/ LBP.
Bendebba et al., 2007	II	Study of 2539 pts w/ LBP who were assessed pre-treatment & 2 yrs following treatment using a new disease-specific outcomes instrument: the Lumbar Spine Outcomes Questionnaire (LSOQ). The reliability ICCs were >0.8 for LBP, leg pain, functional disability, & physical symptoms other than pain. The validity was verified by comparing the results to ODI & SF-36. Responsiveness was verified.	The LSOQ appears to be reliable, valid, & responsive in pts w/ LBP.
Lee et al., 2008	II	Study of 98 pts w/ cervical & lumbar disorders scheduled for surgery comparing the validity of SF-12 (v 2) to SF-36 (v 2). The PCS & MCS scores correlated strongly btwn SF-12 & SF-36 (r range 0.88–0.97). Except for general health, most of the other subscales correlated strongly (r range 0.81–0.99).	The SF-12 (v 2) is a valid alternative to the SF-36 for pts w/ lumbar spine disorders.
Copay et al., 2008	II	454 of 460 pts who underwent lumbar spinal surgery were evaluated to assess the MCID for ODI, SF-36 PCS, & back & leg pain scales. Outcomes instruments were administered preoperatively & 1 yr postoperatively. MCID values: ODI, 12.8 points; SF-36 PCS, 4.9 points; back pain, 1.2 points; leg pain, 1.6 points.	Using 2 different anchors (HTI of the SF-36 & Satisfaction & Results scales), it was possible to calculate MCID values for ODI, SF-36 PCS, & back & leg pain scales.
Guilfoyle et al., 2009	II	620 pts undergoing spinal surgery studied. Patients were evaluated using SF-36, disease-specific instruments, & VAS scores. There was 88% follow-up at 3 mos & 74% follow-up at 12–60 mos. Strong correlations btwn SF-36 physical function & bodily pain domains & specific disability scales were observed. SF-36 physical function, bodily pain, general health, vitality, & mental health domains were free of floor & ceiling effects.	The SF-36 is valid for assessing functional outcomes following lumbar spinal surgery.

(continued)

TABLE 1: Assessment of functional outcome: summary of evidence* *(continued)*

Authors & Year	Level of Evidence	Study Description	Comments
Walsh et al., 2003	II	970 pts w/ spinal disorders w/ complete baseline data & 3-mo follow-up were used to evaluate the responsiveness of the ODI & summary scales of the SF-36. Follow-up rate is not stated. Based on ROC analysis, measures assessing pain were more responsive than those assessing function. The "bodily pain" domain of the SF-36 was the most responsive to worsening symptoms.	The SF-36 is sufficient to measure health status & function for pts w/ back pain. Disease-specific measures (i.e., ODI) might not be necessary when SF-36 is used. Pain scales seem to be the most responsive measures for pts w/ LBP.
Trief et al., 2006	II	115 (72%) of 160 pts from 2 prospective lumbar fusion trials completed preop & 2-yr postop SF-36, ODI, & VAS (pain) assessments. Higher preop MCS scores predicted less back & leg pain after surgery & better postop ODI scores.	Presurgical emotional status (measured using the SF-36 MCS) is a predictor of pain & ODI outcome 2 yrs after lumbar spinal fusion.
Slover et al., 2006	II	3482 pts who recently underwent lumbar spinal surgery completed 3-mo & 1-yr SF-36 & ODI assessments. Follow-up rate is not stated. The average improvement in SF-36 & ODI was smaller in pts w/ psychosocial comorbidities (active compensation, smoking) or medical disorders, including headaches.	SF-36 & ODI are less responsive when significant psychosocial comorbidities or medical comorbidities were present.
Pahl et al., 2006	II	A cross-sectional, observational assessment of 4442 pts w/ spinal disorders was conducted in order to compare the effect of diagnosis on overall health status. Follow-up rate is not stated. Herniated disc w/ radicular pain, lumbar spinal stenosis w/o deformity, degenerative spondylolisthesis, & painful disc degeneration/ spondylosis were all associated w/ negative impact on all 8 subscales of the SF-36.	Younger pts (age <60 yrs) & pts w/ lumbar disc herniation w/ radicular pain had the greatest negative impact on physical health measured using SF-36.
Svensson et al., 2009	II	101 pts were evaluated using the BIS, SF-36, EQ-5D, & ODI, before undergoing surgery. BIS scales showed 80% units more ordered pairs than disordered when compared w/ the other outcomes measures of pain.	The BIS is a valid disease-specific outcome measure for pts w/ back & leg pain. Responsiveness & reliability were not assessed.
Bridwell et al., 2007	III	Multicenter prospective study of 56 pts w/ degenerative lumbar scoliosis to assess the responsiveness of the SRS-22 instrument. Follow-up rate is not stated. The greatest changes observed from preop state to 2 yrs postop were the SRS self-image domain followed by SRS total, SRS pain, & ODI.	The SRS-22 instrument is more responsive than ODI or SF-12 for detecting improved pain, self-image, & function in pts treated surgically for degenerative lumbar scoliosis.

* BIS = Balance Inventory for Spinal Disorders; DRI = Disability Rating Index; FSQ = Functional Status Questionnaire; HR-QOL = health-related quality of life; HTI = Health Transition Item of the SF-36; ICC = intraclass correlation coefficient; LBP = low-back pain; LSOQ = Lumbar Spine Outcomes Questionnaire; MCID = minimum clinically important difference; ODI = Oswestry Disability Index; PCS = physical component summary scale of the SF-36; pt = patient; RMDQ = Roland-Morris Disability Questionnaire; ROC = receiver operating characteristic; SIP = Sickness Impact Profile; SF-12 = 12-Item Short Form Health Survey; SF-36 = 36-Item Short Form Health Survey; SRS-22 = 22-Item Scoliosis Research Society; VAS = visual analog scale.

ety questionnaire (SRS-22), the Balanced Inventory for Spinal Disorders (BIS), and the Lumbar Spine Outcomes Questionnaire (LSOQ) were compared with the ODI and SF-36, since both of these have been shown to be reliable, valid, and responsive for patients with lumbar degenerative diseases who are undergoing lumbar fusion. However, this is not a direct measure of validity.

Finally, a functional outcome instrument must be responsive. The instrument must be able to detect differences in disease severity among populations and should be able to measure the magnitude of treatment effect.

Summary of Literature From Previous Guidelines

Fairbank and colleagues showed that the ODI is a reliable, valid, and responsive measure for detecting changes in low-back pain and its functional severity.[8] Roland and Morris demonstrated the Roland-Morris Disability Questionnaire (RMDQ) is a reliable assessment of acute low-back pain.[18] Deyo showed the Sickness Impact Profile (SIP) and the modified RMDQ are reliable for the assessment of low-back pain, which appears to follow the physical dimension of functional disability.[5] Salén et al. found the Disability Rating Index (DRI) to be a reliable, valid, and responsive measure in patients with axial skeletal pain (see Table 1).[19]

Minimum Clinically Important Difference

The validation of functional outcome measures allows the researcher to confidently select appropriate tools for clinical studies. In order for clinicians to interpret the relevant changes in a particular outcome score, it is important to define the minimum change that is clinically meaningful. Copay and colleagues performed a rigorous study of 460 patients where preoperative and 1-year postoperative scores were obtained in 454 patients with 99% follow-up.[3] The authors determined the MCID for the ODI (12.8 points), SF-36 physical component summary (PCS) (4.9 points), visual analog scale (VAS) for back pain (1.2 points), and VAS for leg pain (1.6 points). The study used robust and validated techniques to provide Level II evidence (see Table 1).[3]

General Health-Related Quality-of-Life Measures

Lee et al. performed a study of 98 patients scheduled for either lumbar or cervical spine surgery and compared the 12-Item Short Form Health Survey (SF-12, version 2) to the SF-36 (version 2).[14] The physical and mental component summary scores strongly correlated between SF-12 and SF-36: r ranged between 0.88 and 0.97. Except for general health, most of the other subscales correlated strongly (r range 0.81–0.99). This study provides Level II evidence that the SF-12 (version 2) is a valid alternative for the SF-36 for patients with lumbar spinal disorders.[14] This is important because of a substantial decrease in the amount of time necessary for eliciting responses on the part of patients by utilization of the SF-12 rather than the SF-36 (see Table 1).

Guilfoyle et al. performed an outcome study of 620 unselected patients who underwent either cervical or lumbar spinal surgery for degenerative disease.[9] The SF-36 was compared with a wide range of disease-specific outcome measures to determine the utility of a general health-related quality of life (HR-QOL) instrument for assessing functional outcome for patients with degenerative spinal diseases. There was excellent early follow-up (88% at 3 months) and a modest loss at long-term follow-up (74% available for follow-up at 1–5 years). The SF-36 physical function, bodily pain, general health, vitality, and mental health domains were free from ceiling or floor effects that would skew the results. In addition, the physical function and bodily pain domains correlated well with validated disease-specific outcome measures. Bodily pain correlated well with VAS arm or leg scores, and the mental health domain correlated well to validated psychological morbidity assessments. The SF-36 physical function and bodily pain domains demonstrated good responsiveness (standard response mean 1.04–1.72 for physical function and bodily pain) following surgery for lumbar disorders. The authors concluded, based on Level II evidence, that the SF-36 was reliable, valid, and responsive for measuring outcome following lumbar spinal surgery (see Table 1).[9]

Walsh et al. assessed outcome at 3 months in 970 patients undergoing a variety of treatments for lumbar degenerative disorders and compared the responsiveness of disease-specific and general health outcome instruments.[23] In this study cohort, 27% of patients underwent surgery, while most were treated with various nonoperative therapies. The authors used a diagnostic test paradigm, the receiver operating characteristic (ROC), for assessing the responsiveness of the different outcome measures. The "gold standard" measure of clinical improvement was physician-patient consensus. Patients did not complete this portion of the assessment 62% of the time, and therefore the level of evidence was downgraded one level for the purposes of establishing recommendations. The bodily pain, physical function, and PCS scores of the SF-36 compared favorably to the ODI. In general, all outcome measures were more responsive for assessing changes in pain than changes in function. The authors provided Level II evidence that the SF-36 is both valid and responsive for assessing lumbar spinal pain and functional outcomes and that it might not be necessary to include disease-specific outcome measures in all studies when using the SF-36 (see Table 1).[23]

Pahl et al. extended the observation that the SF-36 is valid for assessing lumbar spinal disorders by performing a cross-sectional assessment of 4442 patients with spinal problems.[16] The data were generated from the National Spine Network database which consisted of 11,029 patients. The extent of patient follow-up is not stated, and the statistical methods for handling missing data were not discussed. The study's level of evidence was therefore downgraded by one level. These authors found that the impact on patients with lumbar herniated disc with radiculopathy, lumbar stenosis, lumbar degenerative spondylolisthesis, or painful degenerative lumbar spondylotic disc disease was negative in all 8 subscales of the SF-36. Younger patients (< 60 years) and patients with lumbar disc herniation with radiculopathy had the greatest negative impact on physical health as measured by the SF-36.

The authors provided Level II evidence to expand the validity of the SF-36 outcome measure to include patients with lumbar spinal disorders for which surgery is recommended (see Table 1).

Psychosocial Impact on Functional Outcome

Trief et al. explored the effect of a patient's emotional state on functional outcomes following intervention for lumbar spinal disease.[22] In a study comprising 160 patients from 2 separate lumbar fusion prospective trials, the authors obtained follow-up in 115 patients (72%) at 2 years after surgery. They found that the preoperative SF-36 mental component summary (MCS) score was an independent predictor of postoperative ODI score. Specifically, patients with greater emotional morbidity preoperatively had less improvement in ODI following surgery compared with patients with more normal MCS scores (Level II evidence).[22]

Slover et al. made similar observations from a much larger cohort of patients.[20] In a study of 3482 patients who underwent lumbar spinal surgery, the authors found that psychosocial (litigation, chronic headaches, etc.) and medical comorbidities reduced the responsiveness of SF-36 and ODI.[20] The authors' conclusions regarding the effect of psychosocial comorbidities are considered Level II evidence since the rate of follow-up is not stated for this large cohort of patients (see Table 1).

Recently Validated Functional Outcome Measures

It is beyond the scope of the current Guideline Update to provide a comprehensive list of all validated outcomes measures used to evaluate patients with lumbar degenerative diseases. A review of the recent literature, however, did identify 3 relatively novel outcome tools that may prove useful for future outcomes analysis: the Lumbar Spine Outcomes Questionnaire (LSOQ),[1] the Balanced Inventory for Spinal Disorders (BIS),[21] and the Scoliosis Research Society-22 (SRS-22).[2] The LSOQ was found to have an ICC greater than 0.8, was validated by comparing it with the ODI and SF-36 (coefficients of correlation were between 0.7 and 0.9), and was found to be responsive (observed effect sizes ranged from 0.68 to 1.17 for 24-month change scores).[1] These data provide Level II evidence in support of the LSOQ (see Table 1).

The studies evaluating the BIS and SRS-22 were not as comprehensive as those for the LSOQ. The BIS was found to be valid when compared with other outcomes instruments, including the ODI, SF-36, and EQ-5D, but reliability and responsiveness were not reported.[21] The SRS-22 was found to be more responsive than SF-12 or ODI for patients with lumbar degenerative scoliosis who underwent surgical management (Table 1).[2]

Summary

Since the publication of the first generation of lumbar spinal fusion guidelines in 2005, there have been no data that conflict with the previous recommendations. The ODI has emerged as a dominant disease-specific outcome measure. The SF-36 and more recently the SF-12 have emerged as dominant general health outcome measures. In some studies, there are data to suggest that the SF-36 might be sufficient for measuring functional outcome following lumbar spinal fusion because it has demonstrated equivalent responsiveness and validity with disease-specific measures.

More novel outcome measures have been compared with the ODI and the SF-36 to determine their validity and responsiveness. Recent data demonstrate the importance of a patient's pretreatment psychological state as a major independent variable that affects the ability to detect change in functional outcome measures—no surprise to experienced spinal surgeons. Finally, research has established the MCID in major functional outcomes measures, which will enhance the interpretation of these observations. This information will undoubtedly guide future comparative-effectiveness research for lumbar degenerative diseases.

Key Issues for Future Investigation

There is an increasing amount of data suggesting that patient-specific factors, such as pretreatment psychological status, are relevant in the functional outcome assessment following lumbar fusion. Specific diseases are associated with different baseline characteristics that may influence the response depending on the choice of functional outcome measure. The SRS-22, for example, appears to be more responsive than the ODI or the SF-36 for evaluating the results of lumbar spinal fusion in patients with degenerative scoliosis.[2] Establishing whether various functional measures are better suited to assess clinical outcome for a specific degenerative spine disorder will be an important step in the evolution of functional outcome assessment.

Acknowledgments

We would like to acknowledge the AANS/CNS Joint Guidelines Committee (JGC) for their review, comments, and suggestions; Laura Mitchell, CNS Guidelines Project Manager, for her organizational assistance; and Linda O'Dwyer, medical librarian, for assistance with the literature searches. We would also like to acknowledge the following individual JGC members for their contributions throughout the review process: Timothy Ryken, M.D.; Kevin Cockroft, M.D.; Sepideh Amin-Hanjani, M.D.; Steven N. Kalkanis, M.D.; John O'Toole, M.D., M.S.; Steven Casha, M.D., Ph.D.; Aaron Filler, M.D., Ph.D., F.R.C.S.; Daniel Hoh, M.D.; Steven Hwang, M.D.; Todd McCall, M.D.; Jeffrey J. Olson, M.D.; Julie Pilitsis, M.D., Ph.D.; Joshua Rosenow, M.D.; and Christopher Winfree, M.D.

Disclosure

Administrative costs of this project were funded by the Congress of Neurological Surgeons and the Joint Section on Disorders of the Spine and Peripheral Nerves of the American Association of Neurological Surgeons and Congress of Neurological Surgeons. No author received payment or honorarium for time devoted to this project. Dr. Ghogawala receives grants from the Patient Centered Outcomes Research Institute (PCORI) and the National Institutes of Health (NIH). Dr. Groff is a consultant for DePuy Spine and EBI Spine. Dr. Mummaneni owns stock in Spinicity and receives honoraria from DePuy Spine and Globus and royalties from DePuy Spine,

Quality Medical Publishers, and Thieme Publishing. Dr. Wang owns stock in Bone Biologics, AxioMed, Amedica, CoreSpine, Expanding Orthopedics, Pioneer, Syndicom, VG Innovations, PearlDiver, Flexuspine, Axis, FzioMed, Benvenue, Promethean, Nexgen, ElectroCore, and Surgitech and holds patents with and receives royalties from Biomet, Stryker, SeaSpine, Aesculap, Osprey, Amedica, Synthes, and Alphatec. The authors report no other potential conflicts of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author contributions to the study and manuscript preparation include the following. Acquisition of data: all authors. Analysis and interpretation of data: all authors. Drafting the article: Ghogawala. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Ghogawala. Study supervision: Kaiser.

References

1. Bendebba M, Dizerega GS, Long DM: The Lumbar Spine Outcomes Questionnaire: its development and psychometric properties. **Spine J 7:**118–132, 2007
2. Bridwell KH, Berven S, Glassman S, Hamill C, Horton WC III, Lenke LG, et al: Is the SRS-22 instrument responsive to change in adult scoliosis patients having primary spinal deformity surgery? **Spine (Phila Pa 1976) 32:**2220–2225, 2007
3. 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 8:**968–974, 2008
4. Cronbach LJ: Coefficient alpha and the internal structure of tests. **Psychometrika 16:**297–334, 1951
5. Deyo RA: Comparative validity of the sickness impact profile and shorter scales for functional assessment in low-back pain. **Spine (Phila Pa 1976) 11:**951–954, 1986
6. Deyo RA, Andersson G, Bombardier C, Cherkin DC, Keller RB, Lee CK, et al: Outcome measures for studying patients with low back pain. **Spine (Phila Pa 1976) 19 (18 Suppl):** 2032S–2036S, 1994
7. Deyo RA, Diehr P, Patrick DL: Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. **Control Clin Trials 12 (4 Suppl):**142S–158S, 1991
8. Fairbank JC, Couper J, Davies JB, O'Brien JP: The Oswestry low back pain disability questionnaire. **Physiotherapy 66:** 271–273, 1980
9. Guilfoyle MR, Seeley H, Laing RJ: The Short Form 36 health survey in spine disease—validation against condition-specific measures. **Br J Neurosurg 23:**401–405, 2009
10. Hadley MN, Walters BC, Grabb PA, Oyesiku NM, Przybylski GJ, Resnick DK, et al: Guidelines for the management of acute cervical spine and spinal cord injuries. **Clin Neurosurg 49:**407–498, 2002
11. Institute of Medicine: **100 Initial Priority Topics for Comparative Effectiveness Research.** Washington, DC: National Academy Press, 2009 (http://www.iom.edu/~/media/Files/Report%20Files/2009/ComparativeEffectivenessResearchPriorities/Stand%20Alone%20List%20of%20100%20CER%20Priorities%20-%20for%20web.pdf) [Accessed April 17, 2014]
12. Karanicolas PJ, Bhandari M, Kreder H, Moroni A, Richardson M, Walter SD, et al: Evaluating agreement: conducting a reliability study. **J Bone Joint Surg Am 91 (Suppl 3):**99–106, 2009
13. Kopec JA, Esdaile JM: Functional disability scales for back pain. **Spine (Phila Pa 1976) 20:**1943–1949, 1995
14. Lee CE, Browell LM, Jones DL: Measuring health in patients with cervical and lumbosacral spinal disorders: is the 12-item short-form health survey a valid alternative for the 36-item short-form health survey? **Arch Phys Med Rehabil 89:**829–833, 2008
15. Luo X, George ML, Kakouras I, Edwards CL, Pietrobon R, Richardson W, et al: Reliability, validity, and responsiveness of the short form 12-item survey (SF-12) in patients with back pain. **Spine (Phila Pa 1976) 28:**1739–1745, 2003
16. Pahl MA, Brislin B, Boden S, Hilibrand AS, Vaccaro A, Hanscom B, et al: The impact of four common lumbar spine diagnoses upon overall health status. **Spine J 6:**125–130, 2006
17. Resnick DK, Choudhri TF, Dailey AT, Groff MW, Khoo L, Matz PG, et al: Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 2: assessment of functional outcome. **J Neurosurg Spine 2:** 639–646, 2005
18. 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) 8:**141–144, 1983
19. Salén BA, Spangfort EV, Nygren AL, Nordemar R: The Disability Rating Index: an instrument for the assessment of disability in clinical settings. **J Clin Epidemiol 47:**1423–1435, 1994
20. Slover J, Abdu WA, Hanscom B, Weinstein JN: The impact of comorbidities on the change in short-form 36 and oswestry scores following lumbar spine surgery. **Spine (Phila Pa 1976) 31:**1974–1980, 2006
21. Svensson E, Schillberg B, Kling AM, Nyström B: The Balanced Inventory for Spinal Disorders: the validity of a disease specific questionnaire for evaluation of outcomes in patients with various spinal disorders. **Spine (Phila Pa 1976) 34:**1976–1983, 2009
22. Trief PM, Ploutz-Snyder R, Fredrickson BE: Emotional health predicts pain and function after fusion: a prospective multicenter study. **Spine (Phila Pa 1976) 31:**823–830, 2006
23. Walsh TL, Hanscom B, Lurie JD, Weinstein JN: Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. **Spine (Phila Pa 1976) 28:**607–615, 2003

Manuscript submitted March 11, 2014.
Accepted April 1, 2014.
Please include this information when citing this paper: DOI: 10.3171/2014.4.SPINE14258.

Address correspondence to: Michael G. Kaiser, M.D., Columbia University, Neurological Surgery, The Neurological Institute, 710 W. 168th St., New York, NY 10032. email: mgk7@columbia.edu.