

Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 2: assessment of functional outcome

DANIEL K. RESNICK, M.D., TANVIR F. CHOUDHRI, M.D., ANDREW T. DAILEY, M.D.,
MICHAEL W. GROFF, M.D., LARRY KHOO, M.D., PAUL G. MATZ, M.D.,
PRAVEEN MUMMANENI, M.D., WILLIAM C. WATTERS III, M.D., JEFFREY WANG, M.D.,
BEVERLY C. WALTERS, M.D., M.P.H., AND MARK N. HADLEY, M.D.

Department of Neurosurgery, University of Wisconsin, Madison, Wisconsin; Department of Neurosurgery, Mount Sinai Medical School, New York, New York; Department of Neurosurgery, University of Washington, Seattle, Washington; Department of Neurosurgery, Indiana University, Indianapolis, Indiana; Departments of Orthopedic Surgery and Neurosurgery, University of California at Los Angeles, California; Department of Neurosurgery, University of Alabama at Birmingham, Alabama; Department of Neurosurgery, Emory University, Atlanta, Georgia; Bone and Joint Clinic of Houston, Texas; and Department of Neurosurgery, Brown University, Providence, Rhode Island

KEY WORDS • fusion • lumbar spine • practice guidelines • treatment outcome

Recommendations

Standards. It is recommended that functional outcome be measured in patients treated for low-back pain due to degenerative disease of the lumbar spine by using reliable, valid, and responsive scales. Examples of these scales in the low-back pain population include the following: The Spinal Stenosis Survey of Stucki, Waddell–Main Questionnaire, RMDQ, DPQ, QPDS, SIP, Million Scale, LBPR Scale, ODI, the Short Form–12, the JOA system, the CBSQ, and the North American Spine Society Lumbar Spine Outcome Assessment Instrument.

Guidelines. There is insufficient evidence to recommend a guideline for assessment of functional outcome following fusion for lumbar degenerative disease.

Options. Patient satisfaction scales are recommended for use as outcome measures in retrospective case series, where better alternatives are not available. Patient satisfaction scales are not reliable for the assessment of outcome following intervention for low-back pain.

Abbreviations used in this paper: CBSQ = Curtin Back Screening Questionnaire; DPQ = Dallas Pain Questionnaire; DRI = Disability Rating Index; FSQ = Functional Status Questionnaire; JOA = Japanese Orthopaedic Association; LBPR = Low Back Pain Rating; ODI = Oswestry Disability Index; QOL = quality of life; QPDS = Quebec Pain Disability Scale; RMDQ = Roland–Morris Disability Questionnaire; SF-36 = Short-Form–36; SIP = Sickness Impact Profile; VAS = visual analog scale.

Rationale

Lumbar spinal fusion is an increasingly common procedure performed as an adjunct in the surgical management of patients with degenerative lumbar disease and instability. As the frequency and complexity of lumbar fusion surgery increases, there is a tendency for costs and complication rates to increase as well.²⁰ With fewer hospital resources available, the ability to assess objectively the functional outcome following lumbar fusion and to correlate patient outcome with the economic consequences of treatment is important.

Various assessment tools are available for measuring functional outcomes in patients who have undergone lumbar fusion. These outcomes may vary widely in the same population depending on whether subjective or objective measures have been used.¹⁷ Examples of objective outcome measures include physiological, anatomical, economic, health-related QOL, and mortality measurements.¹⁰ Objective outcome measures may be classified into functional questionnaires, global ratings (satisfaction), economic factors (employment, disability, and cost), and physical factors (activities).²¹ The purpose of this review was to identify valid, reliable, and responsive measures of functional outcomes after lumbar fusion for degenerative disease.

Search Criteria

A computerized search of the National Library of Medicine database of the literature published between 1966

and 2003 was performed. A search using the subject heading “lumbar fusion” yielded 3708 citations. The following subject headings were combined: “lumbar fusion and outcomes.” Approximately 204 citations were acquired. Only citations in English were selected. A search of this set of publications with the key words “functional outcome” and “satisfaction” resulted in 107 matches. Alternative searches included each disability index by name. Titles and abstracts of the articles were reviewed and clinical series dealing with adult patients treated with lumbar fusion for degenerative lumbar disease were selected for detailed analysis. Additional references were culled from the reference lists of remaining articles. Among the articles reviewed, 30 studies were included that dealt with lumbar fusion, functional outcomes, and satisfaction surveys. Nineteen of these articles were studies in which the authors examined the reliability of functional outcome measures. In another seven articles investigators examined the utility of these functional outcome measures in the setting of lumbar fusion. Two articles were overviews on functional outcome and lumbar degenerative disease. All papers providing Class I medical evidence are summarized in the evidentiary table (Table 1).

Scientific Foundation

Assessment of Functional Outcome

To assess outcome following treatment properly, a functional instrument must fulfill three criteria.^{11,21} First, it must be reliable.^{10,11} Repetition of the functional assessment should be consistent within (internal reliability) and between (external reliability) observers. If a functional instrument contains multiple domains, each should correlate with the final outcome (internal consistency). Second, a functional instrument must be valid.²¹ It should measure the property intended. For example, an instrument assessing dysfunction due to leg pain would be expected to correlate with a reduction in the ability to walk a given distance. Finally, the instrument should be responsive.²¹ The instrument should be able to detect differences in severity among populations. If an instrument measures low-back pain and this pain improves with physical therapy, the instrument should reflect that improvement quantitatively. When evaluating the utility of a functional tool, the initial assessment should emphasize reliability. If a functional instrument does not produce reliable results, its validity and responsiveness are irrelevant.

In terms of grading the quality of outcomes instruments, κ and α values are used. The κ value refers to the degree of correlation of interrater observations (reliability). In patient-based assessments, it indicates consistency in response at a given time point. The α value, often calculated using the Cronbach α test, reflects the degree to which each domain of a multidomain outcome measure correlates with the final result.⁷ For example, an assessment tool for pain may contain physical, psychological, and social domains. Each domain score should correlate with the final score. For a study to provide Class I medical evidence regarding functional outcomes, the outcomes tool used must have a κ value greater than 0.8. Class II medical evidence requires an outcomes tool to have a κ greater than 0.6. Any outcome scale with a κ value less than 0.6

is considered to provide Class III medical evidence for the assessment of outcomes following an intervention.¹⁸

Roland and Morris^{30,31} followed 230 patients of whom 193 were studied up to 4 weeks after their initial presentation. Functional disability was assessed using a 24-item disability questionnaire (the RMDQ) with statements derived from the SIP and relating to the lower back. Reliability was ascertained in 20 patients with an external reliability greater than 0.91. Internal consistency appeared to be greater than 0.8. Validity was confirmed after comparisons to a six-point pain rating scale and physical signs ascertained by an examining physician.³¹ In this group, 60% of patients appeared to improve over the 4-week period, whereas 20% worsened. Absence from work appeared to correlate less well with disability, as only 8% of the employed were unable to work.³⁰ Using the ODI, Fairbank and colleagues¹² followed 25 patients with acute low-back pain in whom a reasonable prognosis was expected. The questionnaire has 10 categories with six gradations each, for a total score of 50. It was completed at weekly intervals over a period of 3 weeks. Reliability ($\kappa > 0.95$) was confirmed in 22 patients who repeated the questionnaire over 2 days. Validity was demonstrated as patients improved over 3 weeks. Paired t-tests revealed a significant improvement in ODI scores during this time period ($p < 0.005$).

Leclaire and colleagues²⁴ observed patients who presented with acute low-back pain alone (100 cases) or accompanied by radiculopathy (100 cases). The cohort was followed using the RMDQ and ODI questionnaires. In the radiculopathy group, ODI and RMDQ scores were significantly more severe (higher) than in the low-back pain-alone group ($p < 0.0001$). The two scales had a moderate correlation to each other in each subgroup ($r = 0.72$ [radiculopathy]; $r = 0.66$ [lumbago]; $p < 0.0001$). In a cohort of patients with low-back pain, the JOA score was used as a psychometric measure. External reliability was strong ($\kappa > 0.90$) when 15 patients reassessed their status with no change in their symptomatology. Interobserver external reliability among physicians was also sound ($\kappa > 0.90$) in 30 patients reassessed using the JOA. Validity was established by a strong correlation to the RMDQ, ODI, and the SF-36.¹⁵ In several different groups with lumbar degenerative disease, the North American Spine Society Lumbar Spine Outcome Assessment tool was used to assess patients who had undergone conservative or decompressive therapy.⁸ In this study, 136 of 206 questionnaires were successfully completed. External reliability was assessed in 64 patients. Both internal and external reliability was strong ($\kappa > 0.90$). The test was determined to be a valid measure compared with existing instruments.

The SIP is a traditional general functional outcome measure, with 136 items in 12 categories, that has been evaluated in the general populace for a variety of conditions. It has been applied to patients with low-back pain and degenerative lumbar disease. Bergner, et al.,¹ examined the use of this general health instrument in 1108 patients with multiple medical problems including rheumatoid arthritis and hip osteoarthritis.¹ Simultaneous with this questionnaire were a clinician's assessment of physical function and patients' self-assessment of the severity of sickness and dysfunction. In this setting, the test-retest (external) reliability of SIP was greater than 0.90, and its