The assessment of the results of lumbar discectomy is complicated by the nature of the disorder and the goals of the surgery. A herniated lumbar disc is not a life-threatening condition, and lumbar discectomy is most commonly performed to relieve pain or improve a neurological deficit causing functional limitations. Although a clinical improvement may be obvious to the treating physician and the patient, as new procedures and treatments for lumbar disc disease are introduced and tested the relevant question shifts from whether improvement occurred to the degree of that improvement. The latter is more subtle and difficult to determine than the former.

During the past two decades, significant advances have been made in the field of outcome assessment. The primary of a patient’s own evaluation of his/her health is acknowledged among researchers, and instruments to measure these perceptions have been developed and validated. Spine surgeons should have an understanding of the basic methods of outcome assessment and how these developments may be applied to surgery for lumbar disc disease.

TRADITIONAL MEASURES OF OUTCOME

Although “outcomes research” is currently applied to a specific method of patient assessment, authors of papers on lumbar discectomy have reported disability-related symptomatic relief and improvement since the publication by Mixter and Barr in 1934. Particularly in the early stages of development of a new surgical technique, results such as mortality, morbidity, and reoperation rates must be compared with those associated with alternative therapies. Studies in which these factors alone are reported may be considered roughly analogous to Phase I clinical trials that are conducted to examine the safety of medications.

After the safety of a procedure is established, its efficacy relative to alternative operations or nonoperative therapies must be tested. For lumbar discectomy, several ways of measuring patient-related improvement have been used. Improvement assessed using a qualitative scale may be the most common. This method, perhaps best exemplified by the Odom scale, consists of an assessment of the extent of the patient’s improvement. Originally described in relation to outcomes associated with cervical discectomy, the outcome criteria described by Odom as applied to lumbar discectomy are: excellent (“no complaints referable to [lumbar disc] disease” and no functional impairment); good (“intermittent discomfort” without significant functional impairment); satisfactory (subjective improvement but significant functional limitations); and poor (no improvement or worsened condition). The rating is generally performed by an individual other than the patient, often the treating physician. The use of such a scale may be improved by reporting clear criteria for each stratum, but the method is limited by its reliance on an observer’s interpretation of the patient’s clinical status.

Other methods of assessment have been used to reduce interobserver variation and improve objectivity. Data such as LOS and return to work are often relatively easy to ob-
tains from hospital and patient records, even retrospectively. These measures, in addition to being quantifiable, may have the benefit of including an implied economic dimension. A procedure associated with reduced LOS or missed work might be proposed both to improve patient outcome and to improve productivity by decreasing the duration of disability. Unfortunately, the relationship between these factors and health is not clear. Hospital LOS is determined by both the patient and the physician and may be shortened without affecting patient satisfaction and health-related outcome. Time until a patient returns to work is affected by socioeconomic factors such as income and job satisfaction.

Physiological measurements, including straight leg raising and range of motion, may be more objective than observer ratings but may not adequately capture the effect of a lumbar disc disorder on a patient’s life. Symptoms may also be inconsistently present and therefore escape measurement during a short office visit.

Finally, the development of multiaxial noninvasive imaging modalities provides the basis for anatomical outcome measurement. Indices of foraminial decompression may be devised and procedures graded on this or a similar basis. The documented presence of herniated lumbar discs in asymptomatic individuals should serve as a caution against this approach. Conversely, pain and disability may persist despite the fact that radiography and neuroimaging indicate that the operation has been successful.

MEASURING PATIENT OUTCOMES

To address the shortcomings of the aforementioned outcome assessment methods, a variety of outcome questionnaires have been developed. The common goal of these instruments is to measure the patient’s view of his/her health and how it affects his/her ability to perform daily activities. The way in which these instruments achieve this goal may differ, however, and the spine surgeon should understand the different types of outcome questionnaires, their strengths, and shortcomings. Depending on the clinical application or research question, it may be sufficient to use only a single disability questionnaire, or it may be necessary to include several instruments in a protocol to measure the effect of lumbar disc disease and its treatment on different aspects of health.

GENERIC HEALTH-RELATED QOL

Health-related QOL conceptually captures the effects of a patient’s health on his/her overall well-being. Usually, it is thought that HRQOL can be separated into several independent domains or dimensions. These domains, which include emotional well-being, mobility, self-care, sexuality, and social functioning, are individually assessed using HRQOL surveys and are often reported as separate subscores.

Generic HRQOL instruments, which attempt to measure all dimensions of HRQOL and may be applied to a wide variety of populations, are useful for several reasons and should typically be included in effectiveness studies. First, they facilitate comparisons of results among studies. As reporting of generic HRQOL outcomes becomes more common, these comparisons may be between different modalities of therapy (medical compared with surgical) for the same disorder or between treatments of different diseases.

A second benefit of generic HRQOL measurements is that they may reveal unexpected effects of disease on patients’ lives. "Purely" physical disorders may be found to affect emotional functioning. These findings may focus new treatments on previously unrecognized or misunderstood aspects attendant on a disease.

Finally, the literature on several generic HRQOL surveys is extensive. Designing an HRQOL questionnaire is a complex process. Survey instruments must be tested extensively to determine their validity, responsiveness, and reliability. The spine surgeon who is performing or is the consumer of clinical studies or patient outcomes is unlikely to have the time and expertise to design or verify these properties of a new questionnaire. Using established, validated survey instruments eliminates this problem and increases the likelihood that reported results are accurate. Furthermore, there are published studies in which the authors report the magnitude of a change observed in score on these surveys that reflects a clinically significant change observed in a patient with low-back pain or sciatica.

The most commonly used generic HRQOL survey is the SF-36, which contains 36 questions that measure the following eight domains of health: physical functioning, role limitation, social functioning, bodily pain, general mental health, emotional role limitation, vitality, and general health perceptions. Each subscale is scored from 0 to 100, with higher scores indicating better health. Norm-based scoring based on the United States general population transforms the scale so that the mean score for each subscale is 50 and the standard deviation is 10.

PREFERENCE MEASURES

Whereas HRQOL instruments assess an individual’s health status directly, preference surveys attempt to determine the value of each health state as perceived by patients or members of the community. Using questionnaire methods based on game theory, individuals reveal their relative preferences for different health states. The result is a cardinal scale from 0.0 (dead) to 1.0 (perfect health). Although these instruments are not disease specific and may have small effect sizes, they do have at least two advantages. The first is that the scalar for a given health state may be multiplied by the amount of time spent in that state (in years) to calculate quality-adjusted life years. This derived unit may be used to compare interventions with widely different effects on health, particularly those that affect life expectancy and QOL in different ways.

The second advantage of preference or utility measurements is that they reflect the values of the patients or the community. It may be difficult or impossible to determine and measure which aspects of a health state are most important to individuals; preference measurements eliminate the need to question them explicitly about their value system. Incorporating patient or community values into analyses of treatment programs assists both physicians and policymakers in allocating resources appropriately.

The EuroQOL EQ-5D is a commonly used instrument
Lumbar discectomy

for the determination of patient utilities. It consists of five questions and a general-health thermometer scale. Responses to the questions can be combined to define 245 distinct health states. General population–based normative data allow conversion of an individual’s responses to utilities for use in quality-adjusted life years calculations and economic analyses.

DISEASE-SPECIFIC MEASUREMENTS

Because generic HRQOL surveys and preference measures are designed to be applicable to a wide variety of populations with many conditions, they may not capture meaningful differences among patients with a specific condition, such as a herniated lumbar disc. For this purpose disease-specific measurements, which carefully assess disabilities and symptoms commonly caused by a particular disorder, were developed. There are several disease-specific surveys for lumbar spine disorders, including herniated intervertebral disc. A panel of experts has recommended that either the RDQ or the ODI be used in studies of back pain.

The RDQ consists of 24 items concerning functional limitations. Most patients need approximately 5 minutes to complete the survey. The scoring ranges from 0 (minimal disability) to 24 (worst disability). The items were derived from the longer Sickness Impact Profile, which is sensitive to changes in the clinical status of patients with back disorders but, with 136 items, may be too long for routine office use. The phrase “because of my back” was added to each item selected from the Sickness Impact Profile for specificity. The RDQ also has been modified to measure better disability in patients with sciatica by removing two items, adding one item, and including “leg pain” in the wording of several items. This modified RDQ is responsive to changes in clinical status among patients with sciatica over a 3-month period. The difference in RDQ score that corresponds with a clinically significant change may vary between 1 and 2 points and 7 and 8 points, depending on the baseline degree of disability of the cohort. The authors of the modified RDQ recommend a 2- to 3-point difference as the cutoff for a clinically significant change.

The ODI is a 10-item instrument with six response levels per item, in which scores range from 0 (no disability) to 100 (highest disability). This survey also takes approximately 5 minutes to complete. Several modifications to the original instrument have been made to clarify the period of time being assessed. The ODI is valid and responsive to changes in clinical status. A change of between 4 and 15 points is recommended as the cutoff for clinical difference between the mean scores of two groups.

OUTCOMES AFTER LUMBAR DISCECTOMY

Prior to conducting comparative studies between different treatments for lumbar disc disease, the benchmark outcomes for lumbar discectomy must be established. By determining the effectiveness of lumbar discectomy in the routine practice of neurological and orthopedic surgery a standard will be set that any other therapy should meet to be considered a viable alternative treatment. The minimal time required for the completion of most outcome questionnaires allows not only integration into office practice but also adaptation for Internet use. These attributes will facilitate measurement of outcomes of lumbar discectomy in real-world practice (effectiveness) rather than in ideal, unrealistic conditions (efficacy).

The Joint Spine Section of the AANS and CNS, under the guidance of the AANS/CNS Outcomes Committee, conducted a prospective nonrandomized multicenter pilot study of outcomes after lumbar microdiscectomy. Patients were enrolled by the treating physician after consenting to surgery. Short Form–36 and ODI scores were obtained at baseline (preoperatively) and at 6 weeks, 3 months, and 1 year after surgery. Preliminary results were reported for 312 patients, with a retention rate of more than 80%. At 3 months the mean SF-36 physical functioning score was 41.3 points higher than baseline. Increases of 47.6 and 40.4 points were reported for the physical role functioning and bodily pain subscales, respectively, over the same time period. The majority of these improvements persisted at 1-year follow-up examination. A manuscript in which investigators will report the complete results of the entire cohort (> 600 patients) is in preparation.

These preliminary results correlate with those of the Maine Lumbar Spine Study. This study was conducted to determine patients outcome after lumbar discectomy performed by neurosurgeons and orthopedic surgeons relative to those demonstrated in patients who did not undergo surgery. Patients were prospectively nonrandomly enrolled after their treating surgeon determined they were suitable candidates for lumbar discectomy. At 1 year, the mean RDQ score demonstrated in the surgery-related group had improved by more than 11 points, and SF-36 the physical function and bodily pain subscale scores had both improved by more than 40 points.

CONCLUSIONS

Developments in outcomes research have relevance to lumbar disc surgery, particularly because discectomy is expected to improve patients’ QOL by reducing their pain and disability. Valid, responsive, and reliable questionnaires can assist the surgeon in assessing a patient’s response of a patient to treatment. Future studies integrating general HRQOL, patient preference, and disease-specific measurements may help to refine indications for surgery and predict response to nonsurgical or surgical management more accurately. As the time course and pattern of improvement after lumbar discectomy are characterized with sensitive outcome measurement instruments, patient education regarding expected results and recovery will become more detailed and grounded in empirical research. The introduction of new treatments for lumbar disc disease will increase the need for well-designed and executed outcome studies. As demonstrated by the pilot study conducted by the Joint Spine Section of the AANS and CNS, outcomes data for lumbar discectomy can be collected prospectively from multiple centers via the Internet.

References

75 to 80% success rates be the norm? J Neurosurg 96 (Spine 1):34–44, 2002


Manuscript received June 17, 2002. Accepted in final form July 3, 2002.

Doctor Angevine received support from the Wilder Penfield Clinical Investigation Fellowship of the Congress of Neurological Surgeons.

Address reprint requests to: Paul C. McCormick, M.D., M.P.H., Department of Neurological Surgery, Neurological Institute of New York, New York Presbyterian Hospital, 710 West 168th Street, New York, New York 10032. email: pcm6@columbia.edu.