

Evidence-Based Spine Surgery

Daniel K. Resnick, MD

Study Design. Literature review.

Objective. To describe the state of the literature regarding the performance of lumbar fusion for low back pain due to degenerative disease of the spine.

Summary of Background Data. The effectiveness and costs associated with spinal surgery have been a topic of significant debate in both the popular press and professional literature.

Methods. Evidence-based medicine techniques have been applied to many areas of spinal surgery. The results of these analyses are being used by practicing physicians, payors, and others to determine what procedures are appropriate for certain patient populations.

Results. This manuscript describes the methodology, strengths, and weaknesses of evidence-based medicine approaches to spinal surgery. The case for lumbar fusion as a treatment for chronic low back pain due to degenerative disc disease is described as an example.

Conclusion. Evidence-based medicine is a useful tool for summarizing and grading the evidence available in the literature for or against a particular treatment strategy. Its utility is limited by the quality of the primary literature, and the absence of proof cannot be equated with the proof of absence.

Key words: spine surgery, evidence-based medicine, lumbar fusion, outcomes. **Spine 2007;32:S15–S19**

The number of lumbar fusion procedures performed in the United States has increased substantially over the last several years and exhibited an upswing in the late 1990s.¹ There are distinct regional differences in the rate of fusions performed per 1000 patients, a fact that has been interpreted in support of the hypothesis that fusion is overused. Recent editorials in the popular press² and general medical literature¹ have strongly condemned a perceived overutilization of lumbar fusion, and have suggested that the increase in the frequency of fusion surgery noted over the last decade is a result of financial incentives to surgeons and instrumentation companies.² This condemnation is largely based on an apparent lack of evidence to support the role of fusion for the treatment of low back pain. Indeed, Gibson *et al*³ in the 1999 Cochrane review stated “There is no scientific evidence on the effectiveness of any form of surgical decompression or fusion for degenerative lumbar spondylosis compared with natural history, placebo, or conservative management.”

From the Department of Neurological Surgery, University of Wisconsin, Madison, WI.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint requests to Daniel K. Resnick, MD, K4/834 Clinical Science Center, 600 Highland Avenue, Madison, WI 53792; E-mail: Resnick@neurosurg.wisc.edu

■ What Is Evidence-Based Medicine?

Evidence-based medicine is a process by which the literature may be evaluated in a comprehensive and systematic fashion in order to answer clinical questions with varying levels of certainty. In short, evidence-based medicine combines a critical evaluation of the literature with physician judgment and patient values to arrive at a “best guess” solution for a particular problem. When there is high-quality medical evidence to support a particular treatment strategy, then the “best guess” is likely to actually be the best treatment strategy for that particular patient with that particular disorder. If the medical evidence is questionable, then the “best guess” may simply be a treatment strategy that truly represents a well-intentioned and informed guess.

It is possible to make graded recommendations by performing a critical review of the literature to establish the strength of the evidence. Recommendations that are supported by incontrovertible evidence may be recommended with great confidence and labeled as “Grade A,” “Level 1,” or “treatment standards.” Recommendations supported by poor quality or controversial evidence may be reported as “Grade C” or “treatment options.” These types of recommendations are 1 form of clinical practice guidelines. See Table 1 (available online only through ArticlePlus) for a comparison of different grading schemes. Other practice guidelines have been produced that are based on consensus opinion or based purely on economic models. When reviewing a practice guideline, it is important to determine the purpose and methodology used. For example, most of the nursing units at my hospital use “patient care guidelines” written by nurse administrators to ensure that patients receive efficient standardized care. For example, my patients who are treated with lumbar fusion procedures receive a physical therapy consult on postoperative day 1. I am aware of no literature-based evidence to support such a practice, however, it is my opinion that such a practice helps get patients out of bed faster, walking faster, and hopefully home sooner with fewer complications. These guidelines, while valuable, are based solely on local practices and opinions that may not be applicable at other hospitals.

High-quality evidence-based clinical practice guidelines should be transparent. The literature search strategy and strategies used to cull the literature are presented clearly such that if a reviewer wished to recapitulate the review, it would be a straightforward process. Strategies used to evaluate the strength of evidence should be presented clearly, especially when the literature is conflicted or of poor quality. The guidelines should be endorsed by relevant and respected national organizations whose of-

ficers and executive committees have reviewed the documents. The guidelines should be published in a peer-reviewed journal to allow experts on the editorial board to review them. Several such guidelines have been published that directly impact on spine surgery.^{4,5} Of note, guidelines require updating periodically as the literature grows and changes over time.

The main strength of the evidence-based medicine process is the ability to combine a comprehensive literature review with a disciplined and transparent process of evaluation in order to make treatment recommendations in accordance with the strength of the medical evidence. If a treatment strategy is supported by incontrovertible evidence, it is listed as such and is something that should be offered to almost all relevant patients. If a treatment strategy is supported only by poor quality evidence, then the application of that strategy to a given patient should be substantially guided by physician judgment and patient preference. If the government is to require reporting of performance measures, the guideline process can be used to establish the strength of the relationship between that performance measure (*e.g.*, thromboembolism prophylaxis) and patient outcome (*e.g.*, pulmonary embolism) in a particular patient population (*e.g.*, young patient with radiculopathy) undergoing a particular procedure (*e.g.*, outpatient microdiscectomy). If there is no evidence to tie the measure to outcome (as in the example), we are justified in negotiating with the government to eliminate that particular measure. If third-party payors want to know what the evidence is to support the use of a particular treatment strategy, then the guideline process provides a structured summary of the evidence available to support or refute the efficacy of that particular procedure. Finally, if an “expert” witness is making unsubstantiated claims of wrongdoing, then the guidelines can provide a peer-reviewed and nationally endorsed reflection of what truly represents reasonable clinical practice.

■ Limits of Evidence-Based Medicine

There are limits to what evidence-based medicine can address. Literature cannot be interpreted in the absence of common sense and clinical experience. A frequently cited example of the inappropriate application of evidence-based medicine techniques is the assertion that there is no scientific evidence on the effectiveness of parachute use for life preservation following falls from aircraft.⁶ Indeed, no randomized controlled trial (RCT) or even a well-designed cohort comparison has ever been performed to provide such evidence. There are case reports of survival following falls without parachutes and reports of deaths in skydivers using parachutes. A more poignant example of the limitations of literature-based guidelines concerns the evacuation of symptomatic intracranial hematomas. The surgical head injury guidelines recommend removal of such hematomas at an option level⁷ (supported only by low-quality or controversial evidence). In fact, every surgical head injury recommen-

dation is made at an option level. This situation exists simply because no ethical surgeon would withhold available treatment from a patient with a symptomatic hematoma, so no control group exists for comparison. Therefore, just because a treatment is not supported by high-quality medical evidence does not mean that a treatment has no value. Failure to appreciate this conundrum led to the Cochrane review conclusions cited above. The Cochrane group only considers randomized controlled studies as valid sources of medical evidence. The absence of such studies, in their opinion, equates with the absence of evidence.³

■ Evidence-Based Medicine and Lumbar Fusion

Outcome Measures

The application of evidence-based medicine to the lumbar fusion literature was an enormous task. Before reviewers could even begin to evaluate comparative studies, it was necessary to define a number of baseline parameters. The first parameter defined related to outcome measures. A number of commonly used, responsive, validated, and reliable outcome measures are available for the assessment of outcomes following lumbar fusion.^{8–21} The use of such outcome measures allows for a comparison of different treatment strategies. Depending on how the outcome measure is used, however, this may not always be the case. Consider the advantages and disadvantages of minimally incisional technologies for the performance of lumbar fusion. It is widely acknowledged that the main benefits of such procedures are short-term, such as less blood loss and quicker return to work. If a functional outcome measure (even a valid, responsive, and reliable measure) is applied to patients who have undergone fusion procedures 2 or more years following surgery, any potential short-term benefit related to a minimally incisional approach would not be detected.

Similarly, lumbar fusion surgery is not performed on patients with normal lumbar spinal anatomy. In order for a fusion procedure to be contemplated, some evidence of abnormality, usually a form of instability must be demonstrated. Patients with instability of the spine are different from their peers in that they have “bad backs.” Surgical treatment directed at a single level may well provide temporary amelioration of symptoms, however, over time, the strength of this beneficial effect may deteriorate due to the natural history of degenerative spine disease. Therefore a beneficial effect noted 1, 2, or 5 years following surgery may not be present 10 years following surgery. In this case, an outcome study performed 10 years following fusion surgery would fail to demonstrate that surgery had a beneficial effect. Does this truly mean that the patients did not benefit from the procedure?

In order for a study to provide high-quality evidence, in addition to reasonable timing of assessment, the outcome measure used must be valid, responsive, and reliable. For example, when considering patients with low back pain, it is probably reasonable to use an outcome measure that reflects patient pain complaints regarding

the back and legs. It is also reasonable to use a measure of functional outcome to see if the patient has derived improved ability to function within society, take care of himself/herself, continue employment, and enjoy pastimes. Such measures would be considered valid, as they seem to reflect the goals of treatment for surgical intervention. A measure designed to assess psychological stress may or may not be relevant depending on the population studied and interventions offered.

Reliability refers to the ability of the outcome measure to be consistently applied and provide consistent results when employed by the same individual or different individuals to a similar patient population. For example, patient satisfaction scores are notoriously unreliable, as patients will give different answers dependent on the interviewer and situation. Reliability can be measured and is generally reported as a “kappa” value. The higher the value is, the more reliable the measure is. Responsiveness describes the ability of a measure to reflect change that is clinically relevant. In theory, if treatment “A” works better than treatment “B,” then the outcomes achieved with treatment “A” should be measurably better than those achieved with treatment “B.”

Finally, outcome measures appropriate for 1 population undergoing lumbar fusion may not be appropriate for other populations undergoing fusion. For example, return-to-work rate may be a valid, reliable, and responsive outcome measure for patients undergoing anterior lumbar interbody fusion. Is this outcome measure appropriate when we are studying the results of fusion *versus* no fusion in patients undergoing decompression for degenerative spondylolisthesis associated with lumbar stenosis?

Clinically Relevant Benefit (Minimum Clinically Important Difference)

Once we have established a valid, reliable, and responsive outcome measure, we then need to figure out how much of a change is really important. In other words, we need to define a clinically relevant benefit. For example, in a study of the use of bone morphogenetic protein as a substitute for autograft, the authors^{22–24} note a “significant” decrease in blood loss in the bone morphogenetic protein group. The magnitude of this decrease was 66 mL per patient. Measured blood loss is a reliable, valid, and responsive outcome measure. However, is this difference in blood loss a clinically relevant benefit? Does this justify an extra \$5000 per patient? This question is even more relevant when we consider the procedure performed; 1 iliac vein injury in 1 patient could completely change the relationship between the groups with regard to average blood loss.

Study Design and Treatment Effect

Study design may also influence the significance of results. The most common problems relate to sample size. Fritzell *et al.*²⁵ in a randomized series comparing fusion techniques for low back pain, found that there was significant functional improvement in 70% of patients treated with posterolateral fusion with pedicle screws

compared to 60% of those patients treated with posterolateral fusion alone. The study was not designed to detect this level of improvement (see below), and, therefore, this difference in outcomes was not found to be significant. Why is not a 10% absolute improvement in outcome significant? Compare these data to that published in the NASCIS III study, where a very large patient population was studied in order to establish that a small difference in clinical outcome was statistically significant.^{5,26,27}

The strongest medical evidence, here labeled as “Class I medical evidence,” in support for a given treatment is derived from well-designed and appropriately powered randomized controlled clinical trials. If an RCT is poorly designed or underpowered, the quality of the evidence that it provides is downgraded to class II or III medical evidence. When trials of similar quality provide conflicting conclusions, the design of each study is examined closely in order to determine which study is better designed to provide an answer to the clinical question posed. In addition to the issues related to outcome measures and power alluded to above, specifics of trial design may be considered. The CONSORT group²⁸ has published a set of criteria for the grading of clinical trials that allows for a rational application of this principle.

■ Comparison of Two Important Studies

Two randomized controlled clinical trials were identified that described a comparison between the efficacy of surgery (fusion) to nonsurgical management of chronic low back pain due to degenerative disease of the lumbar spine at L4–L5, L5–S1, or both levels. Fritzell *et al.*²⁹ published the results of a multicenter RCT from the Swedish Lumbar Spine Study Group in 2001. These authors assumed that very few patients would improve with conservative care and that a modest proportion of patients treated surgically would improve. They performed a power analysis based on this premise in order to have an 80% power to detect a significant difference in the effect of surgery *versus* the effect of nonsurgical treatment. In other words, they determined how much of an improvement they thought would be clinically relevant and figured out how many patients they needed to include in order to be able to detect that degree of improvement 80% of the time. In this study, 294 patients with disabling back pain who were felt to be surgical candidates were randomized to conservative care (*i.e.*, physical therapy supplemented with education and other pain relieving technologies at the discretion of the treating physician) or 1 of 3 surgical treatment arms. Patients were required to have suffered from back pain for at least 2 years, and to have radiographic and clinical evidence of spondylolisthesis at L4–L5, L5–S1, or both levels. The groups were comparable in all demographic variables measured with the exception of a higher incidence of medical comorbidity in the surgical group. Patients were observed for 2 years with intermediate evaluations at 6 months and 1 year following onset of treatment. Outcomes were assessed using multiple well-validated outcome measures, including pain visual analog scales, the Oswestry Low Back Pain Ques-

tionnaire, the Million Visual Analogue Scale, the General Function Scale, Work Status, and a patient satisfaction survey, and an independent functional assessment by a second spinal surgeon.²⁹

Follow-up was achieved in 98% of patients. Appropriate statistical analysis was performed based on the type of data derived from the different outcome measures. The surgical group did significantly better in terms of pain relief, degree of disability as measured by the Oswestry, Million, and General Function Scale, return-to-work status, and degree of satisfaction reported by the patients and the independent observer. Statistical analysis was rigorous, employing both “intention to treat” as well as a “worst case” scenarios. In short, all primary outcome measures evaluated in the study were significantly improved in the surgical group compared to the nonsurgical group.²⁹ The guidelines group felt that this study provided high-quality evidence demonstrating that lumbar fusion is associated with better outcomes than standard conservative care for appropriately selected patients.

Proponents of various nonsurgical therapies criticized the Fritzell *et al*²⁹ study. For example, Mooney³⁰ commented that the study was unfairly biased against conservative care because the patients had already failed a trial of the same type of therapy before entry in the study. This criticism appears to be valid, given the *a priori* assumptions made by Fritzell *et al*²⁹ in their initial power analysis. This criticism does not, however, diminish the finding that patients treated with lumbar fusion have superior clinical outcomes compared to similar patients treated with usual medical care or those left to suffer the natural history of disabling low back pain.

In 2003, Brox *et al*³¹ conducted a randomized study evaluating the relative efficacy of instrumented posterolateral fusion *versus* a specific protocol of cognitive intervention and physical therapy in a group of 64 patients (27 fusion, 37 nonfusion). The primary outcome measure used was a modified Oswestry Disability Index (modified for the Norwegian population).³² Secondary outcome measures included pain visual analog scales, daily use of medication, General Function Scale, Waddell's Fear Avoidance Belief Questionnaire, and a patient satisfaction score. Physical therapists or rehabilitation physicians assessed outcomes at 1 year following initiation of treatment.

Patients enrolled in the surgical arm were treated with instrumented posterolateral fusion. The patients enrolled in the physiotherapy arm underwent a specifically designed program for patients with low back pain that had been previously demonstrated to be more effective than standard conservative care.³³ This program included cognitive therapy designed to address patient fear behavior as well as supervised physiotherapy averaging 2.5 hours per week for 8 weeks. Because of the intensity of the program, most patients stayed at the treatment center in patient hotels. A home exercise program, physician and therapist consultations, various lessons, group therapy sessions, and participation in peer-led discussion groups followed this intensive course.

Ninety-seven percent of patients were observed for 1 year. Both groups improved significantly from baseline on all outcome measures. The improvement in the primary outcome measure, the modified ODI, in the surgical group was 15.6, and the improvement in the physiotherapy group was 13.3. There were very large confidence intervals noted in this as well as other outcome measures assessed, almost certainly due to sample size. This difference in the degree of improvement between the surgical and physiotherapy group was not found to be significant. The surgical group did do significantly better in terms of relief of lower limb pain, and tended to do better than the physiotherapy group in terms of improvement in back pain, emotional distress, and overall success ratings by both the patient and independent observer. The physiotherapy group scored better fear avoidance activity and work, as well as in fingertip-floor distance. Nonsignificant trends were also seen in favor of the physiotherapy group in terms of the General Function Scale and life satisfaction score.³¹

The authors interpret their findings as demonstrating equivalence between their program of physiotherapy and lumbar fusion. The fact that both groups improved significantly also supports the efficacy of both strategies as superior to the natural history of chronic low back pain treated in a conservative fashion. Given the small size of the study groups and very large confidence intervals reported in the paper, the guidelines group felt that the study was underpowered to detect a difference between 2 relatively effective treatment methods. The use of psychological outcome measures to support nonoperative therapy was also questioned, particularly as only 1 of the groups (nonoperative) had been treated with any psychological counseling. Because of these concerns, the guidelines group downgraded the quality of evidence regarding the effectiveness of lumbar fusion *versus* this specific nonoperative therapy. Based on the demonstrable improvements in both groups, it was felt that this paper provided some support for both treatment methods.

■ Summary of Current Status

There are multiple examples of these types of design flaws in the literature concerning lumbar fusion. Unfortunately for the spine surgeon and the patient with low back pain, these design flaws create the impression that many of the procedures we do are not effective. Third-party payors, politicians, and our patients are demanding justification for the potentially risky and certainly expensive procedures that we are performing on otherwise healthy individuals. There are really no ethical issues preventing the performance of appropriately designed randomized controlled studies to examine the relative efficacy of various fusion procedures to noninstrumented posterolateral fusion in the many subpopulations of patients undergoing fusion for low back pain. The challenge is to determine the right procedure for a given patient population, define a clinically relevant difference in outcome using reliable and valid outcome measures, design a study with adequate power, and perform the


study in an era of burdensome HIPAA regulations and public scrutiny.

■ Improving the Literature

In order to improve our literature, we must design studies that are geared toward answering reasonable questions. We need to study techniques in specific patient populations and compare these techniques to “gold standard” techniques or to the natural history of the disease process. For example, disc arthroplasty represents a new, innovative, and expensive potential alternative to fusion procedures for low back pain. When evaluating the results of disc arthroplasty, it is important that the treatment and control groups be appropriately selected. If the inclusion criteria for entrance into the study are too restrictive, then the relevance of the results of the study to the overall patient population may be questioned. Similarly, if the control group is treated in a way that does not accurately reflect the “gold standard” clinical practice, then the importance of a favorable comparison is diminished. In some cases, a high-quality RCT is not feasible due to patient expectations, difficulties encountered with blinding, and ethical concerns. For example, if a clinician is firmly convinced that procedure “A” is better than procedure “B,” can that physician ethically perform at trial comparing “A” and “B?” Clinical insight combined with expertise in clinical trial design will be required in order to provide high-quality medical evidence to support the procedures we perform to improve the quality of life for our patients.

■ Key Points

- The practice of spinal surgery is under intense scrutiny from the lay press, professional press, and third-party payors.
- Evidence-based methodologies allow for a summary and grading of evidence available in the medical literature.
- Recommendations and shortcomings in the literature identified by evidence-based reviews are fertile areas for further research.
- Randomized controlled studies are the gold standard for medical research, however, their application to surgical patient populations are not always practical or ethical.

 tables

References

1. Deyo RA, Nachemson A, Mirza SK. Spinal-fusion surgery—The case for restraint. *N Engl J Med* 2004;350:722–6.
2. Abelson R, Peterson M. An operation to ease back pain bolsters the bottom line, too. *New York Times*. December 31, 2003.
3. Gibson JN, Grant IC, Waddell G. The Cochrane review of surgery for lumbar disc prolapse and degenerative lumbar spondylosis. *Spine* 1999;24:1820–32.
4. Resnick D, Choudhri TF, Dailey AT, et al. Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 17: Bone growth stimulators and lumbar fusion. *J Neurosurg Spine* 2005;2:737–40.
5. Hadley M, Walters BC, Grabb PA, et al. Guidelines for the management of acute cervical spine and spinal cord injuries. *Clin Neurosurg* 2002;49:407–98.
6. Smith GCS, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: Systematic review of randomised controlled trials. *BMJ* 2003;327:1459–61.
7. Bullock RM, et al. Management of extradural hematoma. In: Ghajar J, ed. *Surgical Management of Traumatic Brain Injury*. New York, NY: Brain Trauma Foundation; 2004.
8. Deyo RA. Comparative validity of the sickness impact profile and shorter scales for functional assessment in low back pain. *Spine* 1986;11:951–4.
9. Deyo RA, Andersson G, Bombardier C, et al. Outcome measures for studying patients with low back pain. *Spine* 1994;19(suppl 18):2032S–6S.
10. Fairbank JC, Couper J, Davies JB, et al. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3.
11. Harper AC, Harper DA, Lambert LJ, et al. Development and validation of the Curtin Back Screening Questionnaire (CBSQ): A discriminative disability measure. *Pain* 1995;60:73–81.
12. Lawlis GF, Cuencas R, Selby D, et al. The development of the Dallas Pain Questionnaire: An assessment of the impact of spinal pain on behavior. *Spine* 1989;14:511–6.
13. Leclaire R, Blier F, Fortin L, et al. 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* 1997;22:68–71.
14. Manniche C, Asmussen K, Lauritsen B, et al. Low back pain rating scale: Validation of a tool for assessment of low back pain. *Pain* 1994;57:317–26.
15. Million R, Hall W, Nilsen KH, et al. Assessment of the progress of the low-back-pain patient. *Spine* 1982;7:204–8.
16. 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* 1983;8:141–4.
17. Ruta DA, Garratt AM, Wardlaw D, et al. Developing a valid and reliable measure of health outcome for patients with low back pain. *Spine* 1994;19:1887–96.
18. Salen BA, Spangfort EV, Nygren AL, et al. The Disability Rating Index: An instrument for the assessment of disability in clinical settings. *J Clin Epidemiol* 1994;47:1423–35.
19. Stucki G, Daltroy L, Liang MH, et al. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. *Spine* 1996;21:796–803.
20. Waddell G, Main CJ. Assessment of severity in low-back disorders. *Spine* 1984;9:204–8.
21. Ware JE, Kosinski M, Keller SD. A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220–33.
22. Burkus JK, Heim SE, Gornet MF, et al. Is INFUSE bone graft superior to autograft bone? An integrated analysis of clinical trials using the LT-CAGE lumbar tapered fusion device. *J Spinal Disord Tech* 2003;16:113–22.
23. Burkus JK, Transfeldt EE, Kitchel SH, et al. Clinical and radiographic outcomes of anterior lumbar interbody fusion using recombinant human bone morphogenetic protein-2. *Spine* 2002;27:2396–408.
24. Burkus JK, Dorchak JD, Sanders DL. Radiographic assessment of interbody fusion using recombinant human bone morphogenetic protein type 2. *Spine* 2003;28:372–7.
25. Fritzell P, Hagg O, Wessberg P, et al. Chronic low back pain and fusion: A comparison of three surgical techniques: A prospective multicenter randomized study from the Swedish lumbar spine study group. *Spine* 2002;27:1131–41.
26. Bracken MB, Holford TR. Neurological and functional status 1 year after acute spinal cord injury: Estimates of functional recovery in National Acute Spinal Cord Injury Study II from results modeled in National Acute Spinal Cord Injury Study III. *J Neurosurg* 2002;96(suppl 3):259–66.
27. Hurlbert RJ. The role of steroids in acute spinal cord injury: An evidence-based analysis. *Spine* 2001;26(suppl 24):S39–46.
28. Moher D, Schulz KF, Altman DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1192–4.
29. Fritzell P, Hagg O, Wessberg P, et al. 2001 Volvo Award Winner in Clinical Studies: Lumbar fusion versus nonsurgical treatment for chronic low back pain: A multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine* 2001;26:2521–32.
30. Mooney V. Point of view. *Spine* 2001;26:2532–3.
31. Brox IJ, Sorensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine* 2003;28:1913–21.
32. Holm I, Friis A, Storheim K. Measuring self reported functional status and pain in patients with chronic low back pain by postal questionnaires: A reliability study. *Spine* 2003;28:828–33.
33. Brox IJ, Friis A, Holm I. Patients with chronic degenerative spinal disease—Can conservative treatment reduce the waiting list for surgery? [in Norwegian] *Tidsskr Nor Løgeforen* 1999;119:1784–7.