Evidence-based Guidelines for the Performance of Lumbar Fusion

Daniel K. Resnick, M.D.

INTRODUCTION

The number of lumbar fusions procedures performed in the United States has increased substantially during 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 press[2] and general medical literature[1] have strongly condemned a perceived overuse of lumbar fusion and have suggested that the increase in the frequency of fusion surgery noted during 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, 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".[3] Third party payors, plaintiff's attorneys, journalists, and politicians have responded to such statements in a predictable fashion.

These claims are certainly troublesome for the spine surgeon. Clinical experience and multiple published series seem to indicate that decompressive procedures are effective for radicular complaints and neurogenic claudication, and that lumbar fusion procedures are an effective treatment for intractable low back pain. This disconnect between the perceptions of the Cochrane reviewers, the lay press, and spinal surgeons requires explanation. A first step in the reconciliation of these disparate points of view is to establish which perceptions are supported by the available literature regarding lumbar spine surgery.

In January 2003, the leadership of the Congress of Neurological Surgeons (CNS) charged the spine section to develop a set of evidence-based guidelines for the performance of lumbar fusion for degenerative disease. The lumbar fusion guidelines initiative, therefore, predates the publication of the Deyo editorial and the New York Times piece. This report describes the process used to develop these guidelines and summarizes some of the more important findings of the guidelines committee of the American Association of Neurological Surgeons (AANS)/CNS Joint Section on Disorders of the Spine.

0148-703/06/5301-0279

BACKGROUND

The process of guidelines development is not new to the spine section. In March 2002, "Guidelines for the Management of Cervical Spine and Spinal Cord Injuries" were published.[24] Eleven neurological and orthopedic spine surgeons were recruited and served as the author committee for the lumbar fusion guidelines (see Acknowledgment). A grant proposal was presented to the executive committee of the spine section, and the spine section leadership agreed to fully fund the guidelines development process. It is important to note that this guideline development process was performed without any industry support whatsoever. When the guidelines were completed, they underwent extensive and repeated peer review. Each guideline was reviewed by the Guidelines committee of the AANS/CNS, by the board of the AANS, by the executive committee of the CNS, and by the clinical practice committee of the North American Spine Society (NASS). After review, modification, and approval by each of these bodies, the guidelines were then subject to peer review by the editorial board of the Journal of Neurosurgery. Each guideline was published as a separate peer-reviewed document, and no financial support was provided for their publication, which occurred in June 2005. The authors of the guidelines are deeply indebted to the board members and executans of the CNS, the AANS, the Spine Section, the members of the NASS clinical practice committee, and the editorial board of the Journal of Neurosurgery, for their support, critique, improvement, and ultimate endorsement of the guidelines.

EVIDENCE-BASED MEDICINE AND THE LOW BACK PAIN PATIENT

The phrase "evidence-based medicine" refers to the practice of medicine based on the best available information in the literature. Evidence-based medicine does not refer to

the "ideal" or "correct" practice of medicine. 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 regarding the effectiveness of parachute use for life preservation after falls from aircraft.[4] A randomized clinical trial simply has not been performed, most likely caused by problems with subject recruitment. Other potential difficulties faced while trying to derive meaningful conclusions from the literature include the pace of technological development and the limitations imposed by the application of standardized outcome measures. Evolving technologies and techniques are, by definition, new and evolving. Therefore, the use of such techniques is not generally supported by high-quality medical evidence. Finally, variables, such as length of follow-up and the choice of outcome measures, vary from study to study, and even the highest quality outcome measures may be improperly applied, limiting the usefulness of a study.[6–19]

Despite the known limitations of the literature, it is important that we examine what information we do have in a comprehensive, systematic, and balanced fashion. Although many questions are left unanswered, information does exist that can be used to support or discourage the use of surgical techniques to treat patients with low back pain. It is only through a thorough examination of what we already know that we can focus on what we need to find out.

METHODOLOGY

The formation of a practice guideline begins with the framing of a clinical question. Developing appropriate questions is not always as straightforward as one might think. For example, a question such as "Is lumbar fusion an effective treatment for low back pain," is not answerable. We were forced to divide such questions into more manageable subtopics. To address this issue, we evaluated the literature regarding lumbar fusion for several different patient populations. It is not difficult to imagine that the literature concerning fusion in elderly patients undergoing fusion after laminectomy for stenosis is probably not relevant to the middle-aged executive with "discogenic" low back pain. Ultimately, 16 topics relevant to the performance of lumbar fusion for degenerative disease of the lumbar spine were selected. It is recognized that this set of 16 questions is by no means comprehensive, and many relevant issues have not been addressed.

Once we settled on the questions, a systematic search of the literature housed at the National Library of Medicine was performed using computerized search engines. The search terms used are listed in each individual guidelines document. Theoretically, if the reader were to use the same search terms and use the same time and language constraints, the reader would obtain the same results as the authors. One of the vagaries encountered during the process is that identical searches performed by coauthors within hours of each other often resulted in slightly different results. The titles and abstracts of each reference were then reviewed by the primary author of an individual topic. Many references were immediately discarded because they were irrelevant, described nonclinical studies, were limited to inappropriate populations, or were case reports or reviews. Once the relevant references were collected, each was pulled, converted (if necessary) to a pdf file, and reviewed in detail. Reference lists from reviewed papers were also used to identify relevant information, as was the collective knowledge base of the author group.

References were sorted according to the level of medical evidence that they provided. Well-designed, appropriately analyzed, randomized controlled trials were rare, but when present contributed Class I medical evidence. The majority of the papers reviewed provided Class III medical evidence. This evidence is generally derived from case series, small comparative cohort studies, and a few randomized trials that were thought to have been poorly executed for one reason or another. There were a number of papers that were judged to provide Class II medical evidence. These were usually larger cohort studies with appropriate outcome measures. Every paper that was judged to provide Class I or Class II evidence was scrutinized closely by the author group, and there were disagreements. When conflicts arose, we resolved them through discussion, a process that was rigorous and time consuming.

Once we had graded the evidence, evidentiary tables were constructed and displayed to the author group using LCD projectors. We then worked to make the leap between levels of evidence and strength of recommendation. This was by far the hardest, most contentious, and most controversial part of the process. For example, let us consider the use of pedicle screws as an adjunct to posterolateral fusion (PLF) performed for the treatment of low back pain. As many of you are aware, the use of pedicle screws is strongly linked to improved arthrodesis rates.[26] As all of you who perform lumbar fusions can attest, and as the literature unequivocally states, the use of pedicle screws increases costs and complication rates associated with PLF.[26] Finally, as many of you realize, and, as the literature also clearly demonstrates, the association between successful arthrodesis and good clinical outcome is tenuous. We, as the guidelines committee, were forced to reconcile these literature-based facts into some sort of cogent clinical recommendation.

The dilemma we faced was how to make a recommendation with any level of conviction when there was high-quality evidence that could be used to recommend for or against the use of pedicle screws. The most obvious and least useful solution would be to make two recommendations at the standard level. Recommendation one would be in favor of the use of pedicle screws to increase fusion rates and recommen-

dation two would be against the use of pedicle screws because of higher complication rates. Ultimately, we decided to make an option level recommendation, because we thought that the use of pedicle screws was supported by inconclusive or controversial medical evidence. The wording of the recommendation was revised at least a dozen times by the committee, and then re-revised after thoughtful peer review. Fortunately, there were only a few instances in which we were forced to reconcile differences between high-quality sources of evidence.

Another significant issue we encountered when trying to interpret the medical literature related to the definition of a clinically relevant outcome measure. For example, in a study of the use of bone morphogenetic proteins (BMP) as a substitute for autograft, the authors note a "significant" decrease in blood loss in the BMP group.[20–22] The magnitude of this decrease was 66 ml per patient. Is this a clinically relevant benefit? Does this justify an extra $5000 per patient? Conversely, Fritzell et al., in a randomized series comparing fusion techniques for low back pain, found that there was significant functional improvement in 70% of patients treated with PLF with pedicle screws compared with 60% of those patients treated with PLF alone.[23] The paper was substantially underpowered to detect this level of improvement, and, therefore, this difference in outcomes was not found to be significant. Is this degree of improvement worthwhile? If so, what does it mean if an underpowered trial failed to demonstrate a significant effect?

RESULTS: FUSION FOR LOW-BACK PAIN

First and foremost, the literature does support the performance of lumbar fusion at one or two levels in properly selected patients with chronic low back pain who have not responded to conservative measures. There have been two randomized controlled clinical trials published that describe a comparison between the efficacy of surgery (fusion) to nonsurgical management of chronic low back pain caused by degenerative disease of the lumbar spine at L4-L5, L5-S1, or both levels. Fritzell et al. published the results of a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group in 2001.[26] 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 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 to be able to detect that degree of improvement 80% of the time). In this study, 294 patients with disabling back pain who were thought to be surgical candidates were randomized to conservative care (physical therapy supplemented with education and other pain-relieving technologies at the discretion of the treating physician), or one of three 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 spondylosis 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 followed for 2 years with intermediate evaluations at 6 months and at 1 year after onset of treatment. Outcomes were assessed using multiple well-validated outcome measures, including pain visual analog scales, the Oswestry Low Back Pain Questionnaire, the Million Visual Analogue Scale, the General Function Scale (GFS), Work Status, and a patient satisfaction survey, and an independent functional assessment by a second spinal surgeon.[26]

Follow-up was achieved in 98% of patients. Appropriate statistical analysis was performed on the basis of 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 GFS, return to work status, and degree of satisfaction reported by the patients and by the independent observer. Statistical analysis was rigorous, using "intention-to-treat" as well as "worst-case" scenarios. In short, all primary outcome measures evaluated in the study were significantly improved in the surgical group compared with the nonsurgical group.[26] This study is, therefore, thought to provide Class I evidence demonstrating that lumbar fusion is associated with better outcomes than standard conservative care for appropriately selected patients.

The Fritzell study was criticized by proponents of various nonsurgical therapies. For example, Mooney 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.[27] This criticism seems to be valid, given the *a priori* assumptions made by the Fritzell group in their initial power analysis. This criticism does not, however, diminish the finding that patients treated with lumbar fusion have superior clinical outcomes compared with 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. conducted a smaller (i.e., less powerful) randomized study evaluating the relative efficacy of instrumented PLF versus a specific protocol of cognitive intervention and physical therapy.[28] The primary outcome measure used was a modified Oswestry Disability Index (ODI; modified for the Norwegian population).[29] Secondary outcome measures included pain visual analog scales, daily use of medication, GFS, Waddel's Fear Avoidance Belief Questionnaire, and a patient satisfaction score. Outcomes

were assessed by physical therapists or rehabilitation physicians at 1 year after initiation of treatment.

Patients enrolled in the surgical arm were treated with instrumented PLF. The patients enrolled in the physiotherapy arm underwent a program specifically designed for patients with low back pain that was thought to be more effective than standard conservative care based on a pilot study performed by the authors.[30] This program included significant cognitive therapy designed to lower patient fear as well as supervised physiotherapy averaging 25 hours per week for 8 weeks. Because of the intensity of the program, most patients stayed at the treatment center in patient hotels. This intensive course was followed by a home program based on the exercises prescribed in the supervised portion. In addition, patients in the physiotherapy group were offered individual consultations, lessons, group therapy sessions, and participation in peer-led discussion groups.

Sixty-four patients were randomized, 37 to surgery and 27 to physiotherapy. There were more men randomized to the surgical group, otherwise the groups were comparable. The 1-year follow-up rate was 97%. 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. The 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 the 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 GFS and life satisfaction score.[28]

The authors interpret their findings as demonstrating equivalent results between their program of physiotherapy and lumbar fusion. Given the small size of the study groups and the very large confidence intervals reported in the paper, the evidence provided by the paper is considered to provide Class III evidence concerning the relative efficacy of fusion versus intensive physiotherapy. The paper does not address the usefulness of fusion as a means to alter the natural history of low back pain and is significantly underpowered to detect any differences between any treatments that are even remotely similar. The relevance of the paper may be further questioned given the intensity of the treatment used in the physiotherapy group. It is doubtful that such a program is available to the vast majority of patients treated for low back pain.

SAMPLE SIZE, CLINICALLY RELEVANT EFFECT, AND PEDICLE SCREWS

The importance of sample size and the definition of "clinically relevant effect" cannot be overstated. A large randomized controlled clinical study may demonstrate a "statistically significant effect" of a treatment modality. If the sample size is large enough, a small difference in outcomes may reach significance. Consider the National Acute Spinal Cord Injury Study (NASCIS) spinal cord injury studies.[31–35] In these studies, large numbers of patients were enrolled and a beneficial effect of methylprednisolone on clinical outcome measured with the American Spinal Injury Association (ASIA) scale was identified (in a subgroup of patients). The magnitude of the improvement was small, however, and the use of methylprednisolone was associated with an increased risk of complications.[31–35] Is the small potential benefit of methylprednisolone use worth the increased risk of complications? Not all clinicians think so.[36–38] Here is where the clinician must make a judgement regarding the clinical importance of a 4-point improvement in the ASIA scale versus an increased risk of sepsis. Conversely, a substantial beneficial effect may not be recognized if sample sizes are too small (*Fig. 31.1*). In the ideal situation, a modest-to-large treatment effect would be detected with moderate sample sizes, allowing the detection of a clinically relevant effect (*Fig. 31.1*).

Previously, we discussed the study by Fritzell et al. that examined the role of lumbar fusion for the treatment of low back pain.[26] These authors performed a power analysis to determine how many patients they would need to include in their study to have a reasonable chance of detecting a significant effect. They assumed that the control patients would do very poorly and that the treated patients would do moderately well. They made several assumptions as to what degree of Oswestry or GFS improvement would be considered relevant and were able to demonstrate a significant effect between the surgical and nonsurgical arms.[26] These same authors then published an analysis of their results within the surgical groups. They compared a noninstrumented PLF group to a PLF supplemented with pedicle screws group to a circumferential fusion group. They found that there were no significant differences between the groups in terms of functional outcomes and that complication rates were higher in the instrumented and circumferential groups.[23]

When one examines the results presented in the Fritzell paper, however, it becomes apparent that the group of patients treated with pedicle screw fixation did score better than the PLF alone group on most of the outcome measures reported, including the Oswestry, GFS, and patient satisfaction surveys. There was a relative 40% increase in the degree of improvement on the Oswestry in the group treated with pedicle screw fixation and an increase in successful outcomes from 60% to 70% (PLF alone versus PLF plus pedicle

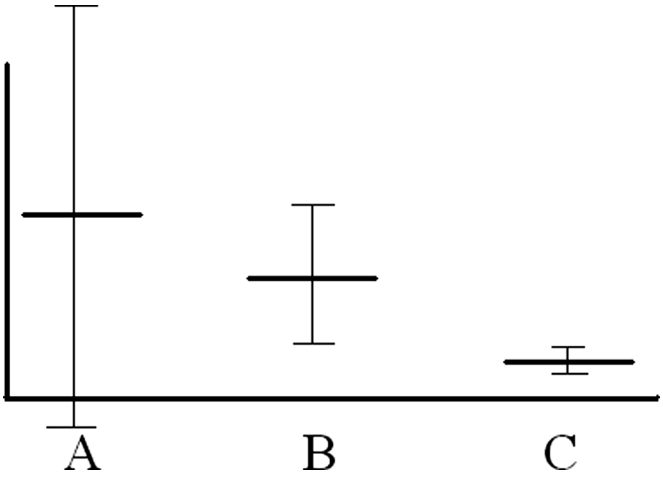

FIGURE 31.1 The deception of power. This graph, adapted from Matthews and Farewell[39] illustrates the problems encountered when trying to interpret studies that are either underpowered or overpowered. On this arbitrary scale, a higher value is associated with a greater beneficial effect. Assume that the heavy line represents the true effect of a treatment. Treatment "A" was studied in a small randomized controlled trial and initially seemed to be very beneficial. Unfortunately, because of a relatively small sample size, there was a large variance within the sample tested. Because of this, a relatively large treatment effect was found to be "nonsignificant." Treatment "C" was studied in a large multicenter study. Because of the large number of patients involved, a very small treatment effect was found to be significant. Therefore, in this example, Treatment "C" would be considered more efficacious that treatment "A," despite the fact that the absolute degree of improvement seen in Treatment "C" was less than that observed in Treatment "A." Treatment "B" was found to have a moderate effect and was detected as significant when studied in an appropriately powered clinical trial.

screws). Is a 40% increase in the degree of improvement on the Oswestry scale or a 16% improvement in rate of good outcomes clinically relevant? If so, why was this difference in outcome not detected as significant?

The problem here is that the Fritzell study was designed to detect a difference between a group of patients who enjoyed a moderate improvement and a group of patients who did not improve much at all. Although the authors were able to detect just such a difference between the surgical and nonsurgical arms, the study was underpowered to detect differences between a group of patients who enjoyed a moderate improvement and a group of patients who had a better improvement. A power analysis reveals that to have an reasonable chance (80%) of detecting a statistically significant difference between a group of patients who achieve a good outcome 60% of the time and another group of patients who achieve a good outcome 70% of the time, more than 350 patients are required in each group (http://calculators.stat.ucla.edu/powercalc). Playing with the numbers, it is possible to calculate that the Fritzell study had only a 42% chance of detecting an effect of this magnitude. Therefore, should we interpret the negative results in the Fritzell study as definitive evidence that the addition of pedicle screws does not improve outcome? The answer is no. The absence of a positive effect in an underpowered study cannot be interpreted as anything except circumstantial evidence (Class III) regarding the lack of a treatment effect.

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 PLF 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 Health Insurance Portability and Accountability Act (HIPAA) regulations and public scrutiny.

CONCLUSION

Despite these limitations of evidence-based literature review, it is imperative that we examine the literature to establish clinical guidelines. These reviews provide snapshots of the state of the literature regarding a particular topic. The guidelines produced are reflections of the peer-reviewed literature and provide valuable guidance regarding what is truly known regarding a particular treatment or diagnostic test. They serve to improve the literature itself through critique and grading of individual papers and through the suggestion of future research directions designed to fill noticeable gaps in our collective knowledge base. These techniques are also being used by agencies outside of medicine to determine which procedures are paid for, which procedures are within the "standard of care," and which devices are approved for use. If we physicians are not intimately familiar with the strengths, weaknesses, and conclusions reached in our own literature, we will forfeit our ability to participate in the formulation of health care policy.

ACKNOWLEDGMENTS

The author acknowledges the other members of the Lumbar Fusion Guidelines Workgroup: Tanvir Choudhri, Andrew Dailey, Michael Groff, Larry Khoo, Paul Matz,

Praveen Mumanneni, Jeff Wang, William Watters, Beverly Walters, and Mark Hadley.

REFERENCES

1. Deyo RA, Nachemson A, Mirza SK: Spinal-fusion surgery: The case for restraint. **N Engl J Med** 350:722–726, 2004.
2. Abelson R, Peterson M: An operation to ease back pain bolsters the bottom line, too. New York Times, New York, 2003.
3. Gibson JN, Waddell G: The Cochrane review of surgery for lumbar disc prolapse and degenerative lumbar spondylosis. **Spine** 24:1820–1832, 1999.
4. Smith GCS, Pell JP: Parachute use to prevent death and major trauma related to gravitational challenge: Systematic review of randomised controlled trials. **BMJ** 327:1459–1461, 2003.
5. Bullock RM, et al.: Mangement of extradural hematoma, in Ghajar J (ed): *Surgical Management of Traumatic Brain Injury*. Brain Trauma Foundation, New York, 2004.
6. Deyo RA: Comparative validity of the sickness impact profile and shorter scales for functional assessment in low back pain. **Spine** 11:951–954, 1986.
7. Deyo RA, Bombardier C, Cherkin DC, Keller RB, Lee CK, Liang MH, Lipscomb B, Shekelle P, Spratt KF, et al.: Outcome measures for studying patients with low back pain. **Spine** 19(Suppl 18):2032S–2036S, 1994.
8. Fairbank JC, Davies JB, O'Brien JP: The Oswestry low back pain disability questionnaire. **Physiotherapy** 66:271–273, 1980.
9. Harper AC, Lambert LJ, de Klerk NH, Andrews HB, Ross FM, Straker LS: Development and validation of the Curtin Back Screening Questionnaire (CBSQ): A discriminative disability measure. **Pain** 60:73–81, 1995.
10. Lawlis GF, Selby D, McCoy CE: The development of the Dallas Pain Questionnaire: An assessment of the impact of spinal pain on behavior. **Spine** 14:511–516, 1989.
11. Leclaire R, Fortin L, Proulx R: A cross-sectional study comparing the Oswestry and Roland-Morris Functional Disability scales in two populations of patients with low back pain of different levels of severity. **Spine** 22:68–71, 1997.
12. Manniche C, Lauritsen B, Vinterberg H, Kreiner S, Jordan A: Low Back Pain Rating scale: Validation of a tool for assessment of low back pain. **Pain** 57:317–326, 1994.
13. Million R, Nilsen KH, Baker RD, Jayson MI: Assessment of the progress of the back-pain patient. **Spine 7**:204–208, 1982.
14. Roland M: 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** 8:141–144, 1983.
15. Ruta DA, Wardlaw D, Russell IT: Developing a valid and reliable measure of health outcome for patients with low back pain. **Spine** 19:1887–1896, 1994.
16. Salen BA, 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.
17. Stucki G, Liang MH, Lipson SJ, Fossel AH, Katz JN: Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. **Spine** 21:796–803, 1996.
18. Waddell G: Assessment of severity in low-back disorders. **Spine** 9:204–208, 1984.
19. Ware JE, Keller SD: A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. **Med Care** 34:220–233, 1996.
20. Burkus JK, 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** 16:113–122, 2003.
21. Burkus JK, Kitchel SH, Watkins RG, Balderston RA: Clinical and radiographic outcomes of anterior lumbar interbody fusion using recombinant human bone morphogenetic protein-2. **Spine** 27:2396–2408, 2002.
22. Burkus JK, Dorchak JD, Sanders DL: Radiographic assessment of interbody fusion using recombinant human bone morphogenetic protein type 2. **Spine** 28:372–327, 2003.
23. Fritzell 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** 27(11):1131–1141, 2002.
24. Hadley M, et al.: Guidelines for the management of acute cervical spine and spinal cord injuries. **Neurosurgery** 50(Suppl):S2–S6, 2002.
25. Moher, D, Schulz, Altman DG: The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. **Lancet** 357:1192–1194, 2001.
26. Fritzell P, Wessberg P, Nordwall A: Swedish Lumbar Spine Study Group: Lumbar fusion versus nonsurgical treatment for chronic low back pain—A multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. **Spine** 26:2521, 2001.
27. Mooney V: Point of view. **Spine** 26(23):2532–2533, 2001.
28. Brox IJ, 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** 28(17):1913–1921, 2003.
29. 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** 28:828–833, 2003.
30. Brox IJ, Friis A, Holm I: Pateints with chronic degenerative spinal disease: Can conservative treatment reduce the waiting list for surgery? **Tidsskr Nor Logeforen** 119:1784–1787, 1999.
31. Bracken MB, et al.: A randomized, controlled trial of methylprednisolone or naloxone in the treatment of acute spinal-cord injury: Results of the Second National Acute Spinal Cord Injury Study. **N Engl J Med** 322:1405–1411, 1990.
32. Bracken MB, et al.: Methylprednisolone or naloxone treatment after acute spinal cord injury: 1-year follow-up data—Results of the second National Acute Spinal Cord Injury Study. **J Neurosurg** 76:23–31, 1992.
33. Bracken MB, et al.: Administration of methylprednisolone for 24 or 48 hours or tirilazad mesylate for 48 hours in the treatment of acute spinal cord injury: Results of the Third National Acute Spinal Cord Injury randomized controlled trial—National Acute Spinal Cord Injury Study. **JAMA** 277:1597–1604, 1997.
34. Bracken MB, et al.: Methylprednisolone or tirilazad mesylate administration after acute spinal cord injury: 1-year follow up—Results of the third National Acute Spinal Cord Injury randomized controlled trial. **J Neurosurg** 89:699–706, 1998.
35. 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** 96(Suppl 3):259–266, 2002.
36. Hurlbert RJ: Methylprednisolone for acute spinal cord injury: An inappropriate standard of care. **J Neurosurg** 93(Suppl 1):1–7, 2000.
37. Hurlbert RJ: The role of steroids in acute spinal cord injury: An evidence-based analysis. **Spine** 26(Suppl 24):S39–46, 2001.
38. Hurlbert RJ, Moulton R: Why do you prescribe methylprednisolone for acute spinal cord injury? A Canadian perspective and a position statement. **Can J Neurol Sci** 29:236–239, 2002.
39. Matthews DE, Farewell VT: *Using and Understanding Medical Statistics*. 3rd ed. Basel, Karger, 1996.

