

MOS Short Form 36 and Oswestry Disability Index outcomes in lumbar fusion: a multicenter experience

Steven Glassman, MD^{a,*}, Matthew F. Gornet, MD^b, Charles Branch, MD^c,
David Polly, Jr., MD^d, John Pelozo, MD^e, James D. Schwender, MD^d, Leah Carreon, MD^a

^aDepartment of Orthopaedic Surgery, University of Louisville School of Medicine and the Kenton D. Leatherman Spine Center,
210 East Gray Street, Suite 900, Louisville, KY 40202, USA

^bStaff Physician at The Orthopaedic Center of St. Louis, 10 Barnes West Drive #200, St. Louis, MO 63141-6350, USA

^cDepartment of Orthopaedics, Wake Forest University Baptist Medical Center, Medical Center Blvd., Winston Salem, NC 27157-1029, USA

^dDepartment of Orthopaedics, University of Minnesota, 2450 Riverside Ave., SR 200, Minneapolis, MN 55454, USA

^eDepartment of Orthopaedic Surgery, University of Texas, Southwestern Medical Center, 8220 Walnut Hill Lane, Suite 101, Dallas, TX 75231-4427, USA

Received 13 April 2005; accepted 3 September 2005

Abstract

BACKGROUND CONTEXT: Patient-based quality of life scales have become a critical element of post-op assessment for lumbar fusion surgery. The most extensive outcomes data have been generated through FDA-regulated IDE trials for new technologies, which produce excellent data but are constrained by strict enrollment criteria and limited indications. This raises a question as to whether the excellent results seen in these IDE trials can be reproduced in standard clinical practice.

PURPOSE: The purpose of this study was to analyze surgical results based upon standardized outcome tools, across a spectrum of interventions, for one- and two-level lumbar spine fusion procedures.

DESIGN/SETTING: This study is a retrospective review of prospectively collected patient based outcomes data.

PATIENT SAMPLE: Four hundred ninety-seven patients, who underwent 1- or 2-level lumbar spine fusion at five participating spine centers, utilizing a variety of surgical techniques. Enrollment criteria included available demographic, surgical and clinical outcome data. At a minimum, patients had pre-op and one year post-op SF-36 data. In many cases two-year post-op SF-36 data and concomitant Oswestry Disability Index (ODI) data was available.

OUTCOME MEASURES: SF36 and ODI.

METHODS: The patient population included 270 females and 227 males, with a mean age of 47 years. Sixty-five percent (N=324) had one level fusions and 35% (N=173) had two level fusions. Demographic data collected included age, gender, BMI, surgical history, smoking history and work status. Data was analyzed with repeated measures analysis of variance (ANOVA).

RESULTS: SF-36 Physical Composite Score (PCS) improved a mean 9.9 points at one year post-op and 9.5 points at two years post-op. ODI improved a mean 22.2 points at one year post-op and 22.1 points at two years post-op. SF-36 PCS data for surgical approach subgroups revealed greater improvement ($p=.03$) in the ALIF group as compared to the PSF, PLIF/TLIF, or 360° fusion groups (12.6 points vs. 8.8, 9.3, 8.4 points) at 1 year post-op. At 2 years post-op, there was greater improvement ($p=.02$) in the ALIF and PSF groups as compared to the PLIF/TLIF and 360° fusion groups (13.8 and 11.2 points vs. 7.7 and 6.3 points). SF-36 PCS data demonstrated similar baseline scores for patients with and without prior decompression, but a significantly greater rate of improvement (11.3 vs. 7.2 points, $p=.002$) for patients without prior lumbar decompression surgery. The ODI data indicated a significantly greater disability at baseline in the prior decompression group, with greater improvement (21.7 vs. 17.5 points) in patients without prior surgery.

FDA device/drug status: not applicable.

Authors acknowledge a financial relationship (research support from Medtronic Sofamor Danek and consultants [SDG, CB, DWP and JP] for Medtronic Sofamor Danek), which may indirectly relate to the subject of this research.

* Corresponding author. 210 East Gray Street, Suite 900, Louisville, KY 40202. Tel.: (502) 584-7525; fax: (502) 584-6851.

E-mail address: tallgeyer@spinemds.com (S.D. Glassman)

CONCLUSIONS: This study documents improved outcomes, based on SF-36 and ODI scores, in patients undergoing lumbar fusion for one and two level degenerative disc disease. The findings also demonstrate efficacy for all of the surgical techniques studied, suggesting that surgeons can appropriately select the surgical strategy with which they are most adept. © 2006 Elsevier Inc. All rights reserved.

Keywords: Lumbar fusion; SF-36; Outcomes data; Oswestry Disability Index

Introduction

Historically, the success of a lumbar spinal fusion surgery has been judged primarily upon whether or not the patient achieves a solid radiographic fusion. More recently, other measures of success, particularly patient-based health status and quality of life scales, have become a critical element of postoperative assessment. The commonly used measures, MOS Short Form 36 (SF-36), Oswestry Disability Index (ODI), and Visual Analog Scale (VAS) are advantageous in that they are standardized, validated, and therefore applicable across similar and dissimilar interventions [1,2].

With regard to lumbar fusion surgery, the most extensive body of patient-based outcomes data has been generated through Food and Drug Administration (FDA) regulated investigational device exemption (IDE) trials for new technologies, including fusion cages, recombinant human bone morphogenetic protein-2 (rh-BMP-2), and artificial disc replacement [3–6]. Although these studies produce excellent data, the patient population is constrained by strict enrollment criteria and limited indications. As an example, enrollment in the rh-BMP-2/LT cage trial was restricted to single-level degenerative disc disease with no greater than Grade 1 spondylolisthesis, no autoimmune disease, and no osteoporosis [7]. This raises a question as to whether the excellent results seen in these IDE trials can be reproduced in standard clinical practice.

In an effort to address these issues, we undertook a multicenter study of standardized outcome measures after lumbar fusion surgery. The participating centers used a variety of surgical techniques, but had in common the collection of prospective patient-based outcome data. The purpose of the study was to analyze surgical results using these standardized outcome tools, across a spectrum of interventions, for one- and two-level lumbar spine fusion procedures.

Materials and methods

This study is a retrospective review of prospectively collected patient-based outcomes data in 497 patients who underwent a primary one- or two-level lumbar spine fusion at five participating spine centers. Enrollment criteria included available demographic, surgical, and clinical outcome data. At a minimum, patients had preoperative and 1-year

postoperative SF-36 data. In many cases 2-year postoperative SF-36 data and concomitant ODI data were available.

The patient population included 270 females and 227 males with an age range of 17 to 86 (mean 47) years. Sixty-five percent (n=324) had one-level fusions, and 35% (n=173) had two-level fusions. One hundred thirty-one patients (26%) had a prior decompressive surgical procedure.

Demographic data collected included age, gender, body mass index (BMI), surgical history, smoking history, and work status. BMI was categorized as underweight, normal weight, overweight, or obese. Surgical data consisted of surgical approach, levels fused, operative time, and estimated blood loss (EBL).

Statistical analysis was performed to assess the overall group in terms of both 1- and 2-year outcome measures. Subgroups based on surgical approach, levels fused, and prior surgery were also evaluated. Data were analyzed with repeated-measures analysis of variance. Between-group comparisons, as well as serial time measurements were made. When appropriate, age, gender, number of surgical levels fused, and BMI were used as covariates, and subsequent statistical comparisons were made on the resulting marginal means. When significant main effects were observed, post hoc analysis was completed with the Least Significant Difference test. P value <.05 was considered significant.

Results

Overall lumbar fusion group

Of the 497 patients treated by lumbar fusion 54% were female and 25% were cigarette smokers. Based on BMI, 25% were normal weight, 37% were overweight, and 30% were obese.

SF-36 and ODI measures showed significant improvement, compared with preoperative outcomes, at both 1 and 2 years postoperatively in all lumbar fusion groups (p=.001). SF-36 Physical Composite Score (PCS) improved a mean 9.9 points (n=497) at 1 year postoperatively and 9.5 points (n=225) at 2 years postoperatively. ODI improved a mean 22.2 points (n=418) at 1 year postoperatively and 22.1 points (n=141) at 2 years postoperatively. There was no statistically significant incremental improvement or deterioration between 1 and 2 years postoperatively

in either measure. There was no differential improvement in SF-36 PCS or ODI based on age, BMI, or smoking status. Slightly better SF-36 PCS scores were observed in women at 1 year postoperatively, but the difference was no longer observed at 2 years postoperatively. Better overall health status, both preoperatively and postoperatively, was noted in younger patients, normal weight patients, and nonsmokers.

Surgical approach subgroups

The surgical procedure performed was an anterior lumbar interbody fusion (ALIF) in 125 patients (25%), posterior or transforaminal lumbar interbody fusion (PLIF/TLIF) in 152 patients (31%), posterolateral spinal fusion (PSF) in 119 patients (24%), and combined anterior and posterior fusion (360° fusion) in 95 patients (19%). Information on surgical procedure was not available on six patients.

Subgroup evaluation by surgical approach was also performed (Table 1). The ALIF group was 48% female, with a mean age of 42 years, and consisted of 80% single-level fusions. The PLIF/TLIF group was 63% female, had a mean age of 55 years, and included 64% single-level fusions. The PSF group was 55% female, with a mean age of 52 years and 70% single-level fusions. The 360° fusion group was 43% female, mean age 44 years, and had 42% single-level fusions. The ALIF and 360° fusion groups were significantly younger ($p < .001$) and included fewer females ($p < .01$) than the PSF or PLIF/TLIF groups. There was a higher incidence of single-level fusions in the ALIF group and two-level fusions in the 360° fusion group ($p < .001$). There were no significant differences between subgroups in terms of BMI, smoking status, or work status. Because there were significant differences in age, gender, and surgical levels between the surgical subgroups, these factors were controlled for in the analysis of outcome measure comparisons between surgical subgroups.

Analysis of the SF-36 PCS data (Figs. 1A, 1B) demonstrates that overall the ALIF patients had better general health status ($p = .002$) and that the 360° fusion patients had poorer general health status both preoperatively and postoperatively ($p = .007$). At 1 year postoperatively, SF-36 PCS data revealed greater improvement ($p = .03$) in the ALIF group compared with the PSF, PLIF/TLIF, or 360° fusion groups (12.6 points vs. 8.8, 9.3, 8.4 points). At 2 years postoperatively, data were available on fewer patients, with 46

Table 1
Demographic characteristics of surgical approach subgroups

Surgical approach subgroup	n	Mean age (years)	Female (%)	Single-level fusion (%)
ALIF	125	42	48	80
PLIF/TLIF	152	55	63	64
PSF	119	52	55	70
360° Fusion	95	44	43	42

ALIF=anterior lumbar interbody fusion; PLIF/TLIF=posterior or transforaminal lumbar interbody fusion; PSF=posterolateral spinal fusion.

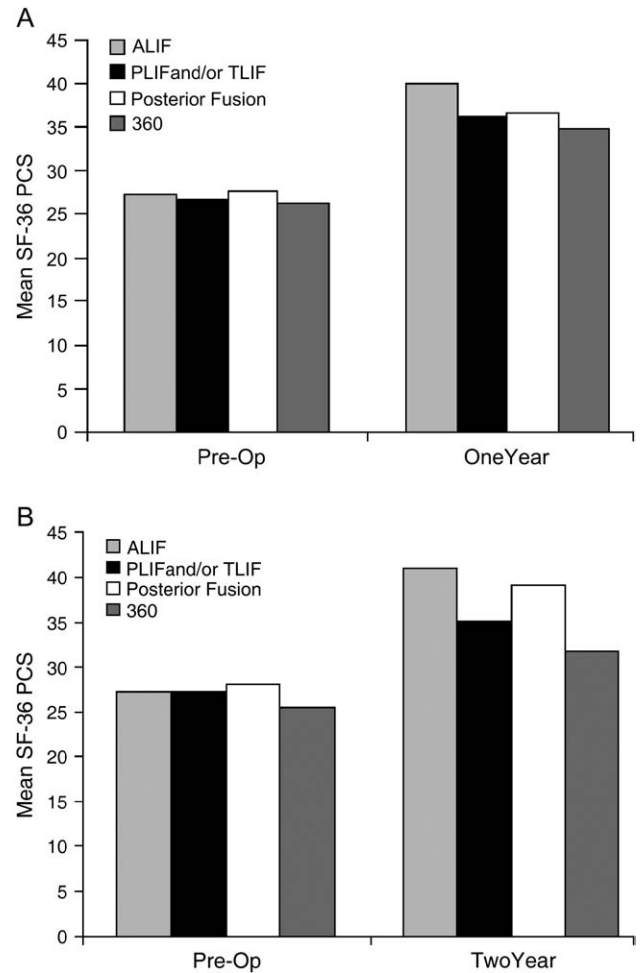


Fig. 1. Bar graphs showing mean SF-36 PCS in each surgical subgroup preoperatively and at one year postoperatively (A); and at two years postoperatively (B). PCS=Physical Composite Score; ALIF=anterior lumbar interbody fusion; PLIF/TLIF=posterior or transforaminal lumbar interbody fusion.

ALIF, 55 PSF, 83 PLIF/TLIF, and 37 360° fusion cases available for analysis. SF-36 PCS showed greater improvement ($p = .02$) in the ALIF and PSF groups as compared with the PLIF/TLIF and 360° fusion groups (13.8 and 11.2 points vs. 7.7 and 6.3 points).

Preoperative ODI scores revealed that PSF patients (mean 55.9) had greater disability than ALIF patients (mean 47.8, $p = .000$), PLIF/TLIF patients (mean 46.1, $p = .000$), or 360° fusion patients (mean 51.4, $p = .023$). At 1 year postoperatively, ODI scores were available in 418 patients (126 ALIF, 119 PSF, 68 PLIF/TLIF, 100 360° fusion). Mean improvement in ODI score was 21.6 points in ALIF patients, 23.1 points in PSF patients, 16 points in PLIF/TLIF patients and 17.9 points in the 360° fusion group. Improvement in ODI varied by surgical approach ($p = .04$) with paired group comparison being statistically significant for the PSF vs. 360° fusion and ALIF vs. PLIF/TLIF groups (Fig. 2).

ODI data were available for only 141 patients at 2 years postoperatively (38 ALIF, 57 PSF, 14 PLIF/TLIF, 31 360°

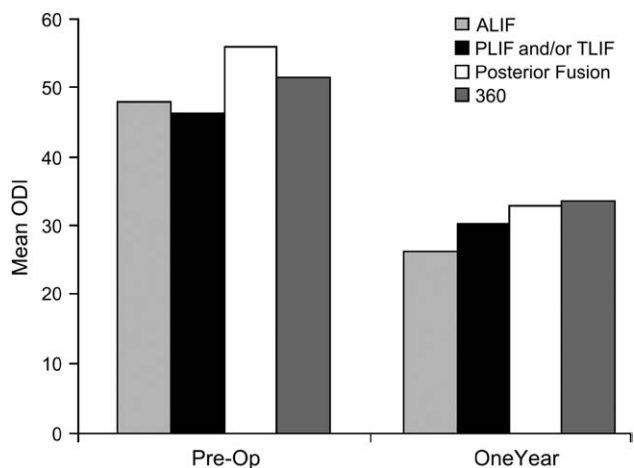


Fig. 2. Bar graphs showing mean ODI scores in each surgical subgroup preoperatively and at two years postoperatively. ODI=Oswestry Disability Index; ALIF=anterior lumbar interbody fusion; PLIF/TLIF=posterior or transforaminal lumbar interbody fusion.

fusion). At 2 years postoperatively, mean ODI improvement was 27 points in the ALIF group, 25.1 points in the PSF group, and 18.9 points in the 360° fusion group. Only 14 PLIF/TLIF scores were available, with mean 4.0 point improvement. The differences between surgical approach groups seen at 1 year postoperatively remained statistically significant ($p=.005$) at 2 year postoperatively.

Surgical levels subgroups

The 324 single-level fusions included 56% females with a mean age of 48.5 years. The 173 two-level fusions were 50% female and a mean age of 49.9 years. There were no statistically significant differences between the groups in terms of smoking history, BMI, or work status. There was no difference in preoperative health status based on either SF-36 PCS or ODI score between one- and two-level fusion patients.

Comparison of preoperative and postoperative outcome measures revealed significant improvement for both groups in SF-36 PCS and ODI at 1- and 2-year intervals (Table 2). There was a trend towards greater improvement for the single-level fusion group on SF-36 PCS at 1 year (1.6 vs. 8.7 points) and 2 years (9.8 vs. 8.9 points) postoperatively. A similar nonsignificant difference between the one- and two-level fusion groups was seen in ODI score at 1 year (21.4 vs. 18.1 points) and 2 years (22.9 vs. 21.0 points) postoperatively.

Table 2
Mean improvement in SF-36 PCS and ODI scores

Surgical levels	1 yr Postop SF-36 PCS	2 yr Postop SF-36 PCS	1 yr Postop ODI	2 yr Postop ODI
One-level	10.6	9.8	21.4	22.9
Two-level	8.7	8.9	18.1	21.0

ODI=Oswestry Disability Index; PCS=Physical Composite Score; SF-36=36-Item Short Form Health Survey.

Prior decompression subgroup

One hundred thirty-one patients had undergone a prior lumbar decompression procedure. The prior decompression group included fewer females (45% vs. 55%). There was no significant difference in patient age (48.5 years vs. 49.2 years), the proportion of smokers (27% vs. 25%), or distribution of BMI. Evaluation of preoperative health status measures revealed no statistically significant difference based on SF-36 PCS (26.1 vs. 27.2, $p=.061$). In contrast, patients with a prior decompression had significantly poorer baseline ODI scores as compared with patients with no prior surgery (mean 54.4 vs. 49.8, $p=.006$).

Assessment of postoperative outcome measures demonstrated a significantly greater improvement in patients with no prior surgery. At 1 year postoperatively, the mean PCS improvement was 11.3 points in the group without prior surgery compared with 7.2 points in the patients with a prior decompression ($p=.000$). A similar difference was observed at 2 years postoperatively (11.6 vs. 6.7 points, $p=.013$). Analysis of ODI data showed that the greater disability seen preoperatively in patients with prior surgery was reproduced at both 1 and 2 years postoperatively. As with the SF-36 PCS data, improvement in ODI at 1 year postoperatively was less in patients with a history of prior surgery (17.5 vs. 21.7 points, $p=.05$). At 2 years postoperatively, the mean improvement was 24.4 points for patients without prior decompression and 16.6 points for patients with prior surgery, but the difference was not statistically significant ($p=.064$) (Fig. 3).

Discussion

Incrementally, prospective patient-based outcome measures have supplanted the physician’s assessment or radiographic parameters as the ultimate gauge of success for lumbar fusion surgery [8,9]. The impetus is multifactorial, but includes the difficulty in effectively determining fusion

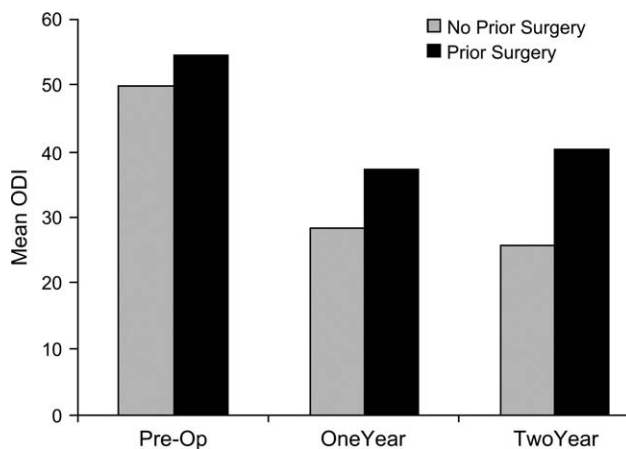


Fig. 3. Bar graphs showing mean ODI scores of patients with prior decompression and those without prior decompression preoperatively and at two years postoperatively. ODI=Oswestry Disability Index.

status [10–13] and the variable correlation between radiographic fusion and clinical outcome [14–16]. Most importantly, standardized measures allow comparison across surgical techniques, treatment paradigms, and even unrelated interventions. As medical and surgical treatments compete for diminishing health-care dollars, it is critical that the value of any intervention be measured from the patient's standpoint in terms of quality of life improvement [17].

A critical element linking societal and individual benefit is the concept of minimal clinically important difference [18,19]. This measure seeks to differentiate a magnitude of change which is not only statistically valid but also of real clinical value. Ware et al. reported that an increase of 5.42 points in SF-36 PCS is clinically important [20]. An older study using ODI indicated that a 4-point decrease might be clinically relevant [21], but a more recent study identified a 10-point decrease as the threshold which patients identify as a significant improvement [22]. FDA standards for a good or excellent outcome include a 15-point improvement in ODI and either maintenance of, or any improvement in, SF-36 PCS score [7]. Ultimately, a compendium of outcome instruments may prove more valuable than any single outcome measure. Although media reports have broadly criticized the effectiveness of lumbar fusion surgery, the largest body of available prospective randomized data appears to contradict those assertions. Outcome data from a series of FDA-regulated IDE trials suggest clinically significant improvement at 2 years postoperatively with an array of lumbar fusion techniques for single-level lumbar degenerative disease. Burkus et al. reported on 254 patients treated with ALIF using LT cages, and documented a mean 29.3-point improvement in ODI and a 13.5-point improvement in SF-36 PCS [3]. In a study of posterolateral instrumented fusion, Dimar et al. reported on 59 patients with a mean improvement of 23.6 points in ODI and 1.9 points in SF-36 PCS [23]. Sasso et al. studied ALIF with cylindrical cages, and these patients had a mean improvement of 21.6 points in ODI and 11.5 points in SF-36 PCS [5]. The control population in that study, femoral ring allograft, had an improvement of 21.2 points in ODI and 8.8 points in SF-36 PCS. The question remains whether the clinical benefit achieved in these well-controlled trials can be sustained in standard clinical practice.

An additional perspective comes from the Swedish Lumbar Spine Study which has generated a series of reports based on a randomized trial of three fusion techniques for L4–L5 and L5–S1 degenerative conditions. At 2 years postoperatively, Fritzell et al. documented ODI improvement of 1.8 points for noninstrumented fusion, 14.8 points for instrumented fusion, and 8.8 points in a combined ALIF/PLIF group [16]. In a comparison to nonsurgical treatment, they reported better results but higher costs with lumbar fusion surgery [8,17]. Important caveats include the use of older surgical technology than the current study, and a potential difference in patient selection, as the authors emphasize the limited use of lumbar fusion in their patient population.

The present study suggests that, to a significant extent, it is possible to generate excellent patient-based health status outcomes in standard clinical practice. Mean improvement of 22.1 points in ODI and 9.5 points in SF-36 PCS compares favorably with prior studies. Although overall improvement in SF-36 PCS and ODI scores was not quite as good as those seen in some of the IDE trials for single-level fusion, the differences were small. Substantial clinical improvement was demonstrated regardless of surgical approach or the number of levels fused. Although only about 50% of the patients had reached 2-year follow-up, clinical benefit was sustained from 1 to 2 years in these patients.

Subgroup analysis within the study was somewhat limited by smaller patient counts and variability in the demographic characteristics of the subgroups. Particularly for the surgical approach subgroups, significant differences in age, gender distribution, and number of levels fused were observed, although statistical correction for these variables was used. Additionally, the surgical indication for lumbar fusion may have differed between the surgical approach subgroups. Given these limitations, outcome measures at 1 year postoperatively were better in the ALIF group as compared with the other three surgical techniques. At 2 years postoperatively, the ALIF and posterolateral fusion groups were equivalent, consistent with prior IDE data on single-level fusions [23]. Both ALIF and PSF subgroups had a greater rate of improvement than the PLIF/TLIF or 360° fusion groups. While differences between surgical technique subgroups were demonstrated, all the surgical techniques generated significant improvement between preoperative and postoperative measures.

Comparison between single-level and two-level fusion procedures was also performed. This revealed a trend towards better outcomes in the single-level subgroup, but no statistically significant differences. The findings were consistent for both SF-36 PCS and ODI scores, at both 1- and 2-year intervals. As opposed to the consistent finding for surgical approach subgroups and the comparisons based on levels fused, the SF-36 and ODI data for patients with a history of prior lumbar decompression yielded potentially conflicting observations.

SF-36 PCS data demonstrated similar baseline scores for patients with and without prior decompression, but a significantly greater rate of improvement for patients without prior lumbar decompression surgery. In contrast, the ODI data indicated a significantly greater disability at baseline in the prior decompression group, with a greater magnitude of improvement in the group without prior surgery. In a cross-sectional study of over 18,000 patients from the National Spine Network, Hee et al. noted that patients with prior lumbar surgery had poorer general health status than unoperated patients [24]. In the present study, the mean baseline SF-36 PCS scores in patients with and without prior surgery (26.1 points and 27.2 points) were similar to the mean 27.4 point PCS score reported for patients with prior surgery in Hee's study. This difference may be explained

by the fact that our entire study population was a surgical cohort.

The strengths and weaknesses of this study both stem from the multicenter design. This approach affords a large volume of cases and therefore statistical power in the analysis. The multicenter approach also generates a realistic view of how a given surgical technique might perform among a range of surgeons. The downsides include small variations in data collection, and a failure to more specifically delineate surgical indications. It seems intuitively clear that different surgical strategies would be best suited to specific pathologies. Degenerative disc disease is obviously too broad a diagnostic entity to support a focused analysis. An important future goal is to undertake a prospective multicenter study in which the surgical pathology is defined more narrowly and consistently across participating centers.

It should be emphasized that this study provides no comparison between surgical and nonsurgical treatment. Despite the obvious importance of such a comparison, study design remains a major challenge. Therefore, the ability to investigate outcomes of surgical versus nonsurgical treatment, for an accurately defined pathology, is another important long-term goal.

Despite these considerations, this study strongly supports the conclusion that lumbar spine fusion is an effective intervention for one- and two-level degenerative disc disease. The observations are substantiated by prospectively collected patient-based outcome measures in a large study population. The findings also demonstrate efficacy for all of the surgical techniques studied, suggesting that surgeons can appropriately select the surgical strategy with which they are most adept. Further investigation is necessary in an effort to optimize surgical strategy for a specific surgical pathology.

Acknowledgments

The authors thank Michael Shainline for his assistance in data analysis and the Lumbar Spine Study Group for their support.

References

- [1] Walsh TL, Hanscom B, Lurie JD, Weinstein JN. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability index, MODEMS and the SF-36. *Spine* 2003;28:607–15.
- [2] Fairbank JCT, Pynsent PB. The Oswestry Disability Index. *Spine* 2000;25:2940–53.
- [3] Burkus JK, Gornet MF, Dickman CA, Zdeblick TA. Anterior lumbar interbody fusion using rhBMP-2 with tapered interbody cages. *J Spinal Disord Techniques* 2002;15:337–49.
- [4] Burkus JK, Heim SE, Gornet MF, Zdeblick TA. 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 Techniques* 2003;16:113–22.
- [5] Sasso RC, Kitchel SH, Dawson EG. A prospective, randomized controlled clinical trial of anterior lumbar interbody fusion using a titanium cylindrical threaded fusion device. *Spine* 2004;29:113–22.
- [6] McAfee PC, Fedder IL, Saiedy S, Shucosky EM, Cunningham BW. SB Charite disc replacement. Report of 60 prospective randomized cases in the US Center. *J Spinal Disord Techniques* 2003;16:424–33.
- [7] FDA Protocol G000137. A prospective, randomized clinical investigation of recombinant human bone morphogenetic protein-2 and compression resistant matrix with the CD Horizon spinal system for posterolateral lumbar fusion in patients with symptomatic degenerative disc disease. Medtronic Sofamor Danek, Memphis, Tennessee, Feb. 2002.
- [8] Fritzell P, Hagg O, Wessberg P, Nordwall A. 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.
- [9] Pellise F, Vidal X, Hernandez A, Cedraschi C, Bago J, Villanueva C. Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. *Spine* 2005;30:365–8.
- [10] Cook SD, Patron LP, Christakis PM, Bailey KJ, Banta C, Glazer PA. Comparison of methods for determining the presence and extent of anterior lumbar interbody fusion. *Spine* 2004;29:1118–23.
- [11] Blumenthal SL, Gill K. Can lumbar spine radiographs accurately determine fusion in postoperative patients? Correlation of routine radiographs with a second surgical look at lumbar fusion. *Spine* 1993;18:1186–9.
- [12] Siambanes D, Mather S. Comparison of plain radiographs and CT scans in instrumented posterior lumbar interbody fusion. *Clin Orthop* 1998;21:165–7.
- [13] Christensen FB, Laursen M, Gelineck J, Eiskjaer SP, Thomsen K, Bungler CE. Interobserver and intraobserver agreement of radiograph interpretation with and without pedicle screw implants. *Spine* 2001;26:538–44.
- [14] Fischgrund JS, Mackay M, Herkowitz HN, Brower R, Montgomery DM, Kurz LT. 1997 Volvo award winner in clinical studies. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal instrumentation. *Spine* 1997;22:2807–12.
- [15] Moller H, Hedlund R. Instrumented and noninstrumented posterolateral fusion in adult spondylolisthesis. *Spine* 2000;25:1716–21.
- [16] Fritzell P, Hagg O, Wessberg P, Nordwall A. 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.
- [17] Fritzell P, Hagg O, Jonsson D, Nordwall A. Cost-effectiveness of lumbar fusion and nonsurgical treatment for chronic low back pain in the Swedish lumbar spine study. A multicenter, randomized, controlled trial from the Swedish lumbar spine study group. *Spine* 2004;29:421–34.
- [18] Farrar JT, Portenoy RK, Berlin JA, et al. Defining the clinically important difference in pain outcome measures. *Pain* 2000;88:287–94.
- [19] Bolton JE. Sensitivity and specificity of outcome measures in patients with neck pain: detecting clinically significant improvement. *Spine* 2004;29:2410–7.
- [20] Ware JE, Kosinski M, Keller SK. SF-36 physical and mental health summary scales: a user's manual. Boston, MA: The Health Institute, 1994.
- [21] Meade TW, Dyer S, Browne W, et al. Low back pain of mechanical origin: randomized comparison of chiropractic and hospital outpatient treatment. *BMJ* 1990;300:1431–7.
- [22] Hagg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003;12:12–2.
- [23] Dimar JD, Glassman SD, Burkus JK, Carreon LY. A comparison of prospective clinical outcome measures for anterior vs posterior single-level spine fusion. *Spine J* 2004;4:109–11.
- [24] Hee HT, Whitecloud TS, Myers L. The effect of previous low back surgery on general health status: results from the National Spine Network initial visit survey of patients with low back pain. *Spine* 2004;29:1931–7.