Comparative Effectiveness Evidence from the Spine Patient Outcomes Research Trial

Surgical Versus Nonoperative Care for Spinal Stenosis, Degenerative Spondylolisthesis, and Intervertebral Disc Herniation

Anna N. A. Tosteson, ScD,* Tor D. Tosteson, ScD,* Jon D. Lurie, MD, MS,* William Abdu, MD, MS,* Harry Herkowitz, MD, † Gunnar Andersson, MD, PhD, ¶ Todd Albert, MD, § Keith Bridwell, MD, ¶ Wenyan Zhao, MS,* Margaret R. Grove, MS,* Milton C. Weinstein, PhD, || and James N. Weinstein, DO, MS*

Study Design. Cost-effectiveness analysis of a randomized plus observational cohort trial.

Objective. Analyze cost-effectiveness of Spine Patient Outcomes Research Trial data over 4 years comparing surgery with nonoperative care for three common diagnoses: spinal stenosis (SPS), degenerative spondylolisthesis (DS), and intervertebral disc herniation (IDH).

Summary of Background Data. Spine surgery rates continue to rise in the United States, but the safety and economic value of these procedures remain uncertain.

Methods. Patients with image-confirmed diagnoses were followed in randomized or observational data with data on resource use, productivity, and EuroQol EQ-5D health state values measured at 6 weeks, 3, 6, 12, 24, 36, and 48 months. For each diagnosis, cost per quality-adjusted life year (QALY) gained in 2004 US dollars was estimated for surgery relative to nonoperative care using a societal perspective, with costs and QALYs discounted at 3% per year.

Results. Surgery was performed initially or during the 4-year follow-up among 414 of 634 (65.5%) SPS, 391 of 601 (65.1%) DS, and 789 of 1192 (66.2%) IDH patients. Surgery improved health, with persistent QALY differences observed through 4 years (SPS QALY gain 0.22; 95% confidence interval: 0.15, 0.34; DS QALY gain 0.34, 95% CI: 0.30, 0.47; and IDH QALY gain 0.34, 95% CI: 0.31, 0.38). Costs per QALY gained decreased for SPS from $77,600 at 2 years to $59,400 (95% CI: $37,059, $125,162) at 4 years, for DS from $115,600 to $64,300 per QALY (95% CI: $32,864, $83,117), and for IDH from $34,355 to $20,600 per QALY (95% CI: $4,539, $33,088).

Conclusion. Comparative effectiveness evidence for clearly defined diagnostic groups from Spine Patient Outcomes Research Trial shows good value for surgery compared with nonoperative care over 4 years.

Key words: spinal stenosis, degenerative spondylolisthesis, intervertebral disc herniation, fusion surgery, instrumented fusion, Cost-effectiveness, EQ-5D, SF-6D, QALY. Spine 2011;36:2061–2068

The American Recovery and Reinvestment Act of 2009 mandated a $1.1 billion investment in comparative effectiveness research, defined by the Institute of Medicine as “…the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition…..”1 Although the role of economic endpoints in comparative effectiveness research remains controversial, the marked growth in complex spine surgery and accompanying expenditures in the US population over the past two decades has prompted concern regarding spine surgery’s value for both individual patients and society.2–4 Begun more than a decade ago, the Spine Patient Outcomes Research Trial (SPORT) addresses Institute of Medicine priority conditions and the comparative effectiveness of surgery and nonoperative care, using clinical and economic endpoints from both randomized and observational study cohorts.5–10

SPORT was designed with a secondary objective of assessing the cost-effectiveness of spine surgery for patients with back and/or leg symptoms for three specific clinical conditions. The economic value of surgery relative to nonoperative care at 2 years compared favorably with many health interventions.11,12 However, surgery for degenerative spondylolisthesis (DS) was somewhat more costly than for patients with stenosis alone (mean cost per quality-adjusted life year was $33,088 for surgery and $20,600 for nonoperative care).12

www.spinejournal.com 2061
[QALY] gained $115,600 vs. $77,600 for stenosis alone). This was largely due to differences in the initial cost of surgery for patients with DS; these patients often undergo fusion surgery, which is more costly than decompressive laminectomy alone (the most common procedure in patients who have only stenosis). In contrast to prior literature, we hypothesized that surgery’s value—in these well-defined conditions—would improve over time. This would occur if health gains remained durable, especially if patients receiving surgery had lower ongoing health care costs relative to nonoperatively treated patients, taking into account the offsetting cost of repeat surgeries, which would have the potential to diminish surgery’s cost-effectiveness. In this article, we report SPORT 4-year cost-effectiveness outcomes for all patient groups.

MATERIALS AND METHODS
Details of SPORT’s design and conduct are provided elsewhere.6–8,10,11 In brief, participants enrolled in either a randomized or observational cohort from 13 participating US multidisciplinary spine practices in 11 states between March 2000 and March 2005 and were followed for outcomes over 4 years. Participants in the randomized group were assigned treatment whereas those in the observational cohort chose their treatment. Eligible participants were aged 18 years and older with well-defined symptoms, physical findings, and imaging-confirmed diagnosis of spinal stenosis (SPS) either alone or associated with DS, or diagnosis of intervertebral disc herniation (IDH). Nonoperative treatments were “usual care” determined by patients’ and physicians’ choice. For SPS, the protocol surgical intervention was a standard posterior laminectomy. For DS, the protocol surgery was the same procedure with or without bilateral single-level fusion with or without instrumentation. For IDH, the protocol surgical intervention was a standard open discectomy. An independent Data Safety and Monitoring Board oversaw the study and a human subjects committee approved the protocol at each institution.

Treatment Effectiveness
For the cost-effectiveness analysis, treatment effectiveness was measured using QALYs, which account for both length and quality of life,14 by weighting time spent in each health state by a health state value. Health state values—reflecting societal health preferences on a scale where a year in best imaginable health is assigned a value of 1 and death is assigned a value of 0—were obtained using the EuroQol instrument (EQ-5D) with US scoring.15,16 Secondary analyses used the short-form (SF)-6D (UK scoring) health state values derived from SF-36 health status instrument.17 Mean health states were estimated at baseline, 6 weeks, 3, 6, 12, 24, 36, and 48 months.

Treatment Cost
Participants were given health care diaries to assist them in tracking both medical resource use and time lost from work and other activities. Total costs included direct medical costs (on the basis of patient-reported utilization; limited to spine-related services except for physician visits, and hospitalizations) and indirect costs (on the basis of patient-reported time away from work and/or usual activities because of spine-related problem[s]). Information was collected from patients with questionnaires at each time point, using either a 6-week (at 6 weeks and 3 months) or 1-month recall period. Reports of hospitalizations, surgeries, and device use were not confined to a recall window.

Direct medical costs included any emergency department or outpatient visits (surgeons, chiropractors, other physicians, physical therapists, acupuncturists, or other health care providers) and spine-related diagnostic tests (radiograph, computed tomographic scan, or magnetic resonance imaging); electromyography; injections; devices (e.g., braces, canes, and walkers); medications; and rehabilitation or nursing home days. To estimate direct medical costs, unit costs were assigned to each visit, test, and procedure on the basis of 2004 Medicare national allowable payment amounts18, with medication costs based on 2004 average wholesale prices.19 For each participant, medical resource use was multiplied by unit costs to estimate total direct medical cost at each time point. All costs are expressed in 2004 US dollars.

Surgery costs depended on the procedure performed and occurrence of complications, which in turn determined the diagnosis-related group. The observed 2004 Medicare mean total diagnosis-related group payment was used to reflect hospital-related surgery costs. Surgeon costs were based on 2004 Medicare allowable amounts, using the resource-based relative value scale.20 Anesthesiology costs were estimated using operative time. For hospitalizations not associated with a spine surgery, costs were based on the diagnosis-related group, using mean observed 2004 Medicare payments.

At each follow-up, the impact of spine-related problems on productivity was assessed. Participants were asked to report missed workdays if employed outside of the home and missed homemaking days if housekeeping was designated as the primary work activity. Use of unpaid caregivers for spine-related problems (including spousal care giving) was also obtained. Costs were estimated using the standard human capital approach by multiplying the change in hours worked by the gross-of-tax wage rate on the basis of self-reported wages at study entry.21 Costs for missed days of housekeeping and unpaid caregivers were valued on the basis of average wages plus nonhealth benefits for individuals aged 35 years and older.22–24

Statistical Analysis
Data were analyzed separately by disease group according to treatment received for the pooled SPORT randomized and observational cohorts, using longitudinal regression models fitted with generalized estimating equations.25,26 Separate models were fit for EQ-5D and for 30-day costs measured at each follow-up time point after surgery or the beginning of nonoperative therapy. If a visit was missing, all other available visits for that patient were included in the analysis.

The treatment indicator (surgery vs. nonoperative care) was a time-dependent covariate, allowing for variable surgery times. After surgery, outcomes were assigned to the
surgical group, with follow-up times measured from the date of surgery. To adjust for potential confounding in each model and the possible effects of missing data, baseline variables associated with missing data or treatment received were included as covariates. All models included a fixed effect for center. To account for correlations among repeated measurements for individuals, including observations before and after surgery, the longitudinal regression models were fit with PROC GENMOD (SAS version 9.1 Windows XP Pro, Cary, NC), specifying a compound symmetry assumption for the working covariance matrix.

Cost-Effectiveness Analysis

The primary cost-effectiveness end point was the incremental cost-effectiveness ratio estimated as cost per QALY gained for surgery relative to nonoperative treatment. For stenosis patients with or without DS, we report cost per QALY gained by surgery type relative to nonoperative care.

Mean total costs and QALYs from baseline to 4 years were estimated for each diagnosis and treatment group using a 3% annualized discount rate for both end points. Discounting is used to weigh near-term costs and health more heavily in the analysis than those occurring in the future. A time-weighted average was used to estimate the difference in QALYs between the surgical and nonoperative treatments on the basis of adjusted mean differences in EQ-5D estimated from longitudinal regression models at each follow-up. QALY differences between treatment groups were estimated using a common baseline EQ-5D value. For costs, mean differences were based on adjusted mean costs summed across time points for each treatment group. To estimate a confidence interval (CI) for the cost per QALY gained, a bootstrap method was applied using 1000 samples taken with replacement from the original sample with the individual as the unit of observation.

Sensitivity analyses of analytic assumptions included restricting analyses to the randomized or observational cohort; limiting costs to direct medical costs only to facilitate comparisons between procedures within disease groups impractical. Among those with SPS, fusion surgery’s cost per QALY gained relative to nonoperative care was $257,600 with a very wide CI (Table 2). Among those with DS, fusion surgery’s cost per QALY gained relative to nonoperative care was $66,300.

When type of instrumentation was examined for DS patients who underwent instrumented fusion, no statistically significant differences in QALY outcomes were found. The cost-effectiveness of each type of instrumentation relative to nonoperative treatment was comparable at approximately $65,000 to $75,000 per QALY gained.

In sensitivity analyses, mortality adjustment, method of QALY estimation, and limiting the analysis to surgeries occurring within 2 years had little impact on cost-effectiveness estimates (Table 3). Although study cohort (randomized vs. not) had little impact on cost-effectiveness for DS or IDH, in the SPS group the randomized cohort cost per QALY gained was somewhat higher at $124,700. Estimates remained less than $125,000 per QALY gained across disease groups when higher surgery costs were used.

DISCUSSION

We used longitudinal patient-reported data on resource use, productivity loss, and health-related quality of life to evaluate...
the cost-effectiveness of surgery relative to nonoperative care for three well-defined clinical cohorts. Compared with findings over 2 years, when assessed over 4 years the value of surgery improved for all groups, and most notably for individuals with DS. This finding warrants examination of both the effectiveness and cost sides of the cost-effectiveness equation in comparison to previously reported 2-year outcomes.11,12

QALY differences at 2 years between surgically and nonoperatively treated individuals of 0.17 (95% CI: 0.12, 0.22) for SPS, 0.23 (95% CI: 0.19, 0.27) for DS, and 0.21 (95% CI: 0.16, 0.25) for IDH were previously reported.11,12 Using these 2-year differences as benchmarks, the 4-year QALY results reported here continue to favor surgery (additional QALY differences of 0.05 for SPS, 0.13 for both DS and IDH for 4-year outcomes minus 2-year outcomes). However, the magnitude of the difference for SPS patients diminished far more than can be explained by the 3% per year discount rate employed in our analysis. For SPS patients, differences favoring surgery

<table>
<thead>
<tr>
<th>TABLE 1. Baseline Participant Characteristics by Disease Group and Treatment Received Within 4 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mean age (SD)</td>
</tr>
<tr>
<td>Women, n (%)</td>
</tr>
<tr>
<td>Ethnicity: not Hispanic, n (%)</td>
</tr>
<tr>
<td>Race, white, n (%)</td>
</tr>
<tr>
<td>At least some college education, n (%)</td>
</tr>
<tr>
<td>Annual income &lt; $50,000, n (%)</td>
</tr>
<tr>
<td>Married, n (%)</td>
</tr>
<tr>
<td>Work status</td>
</tr>
<tr>
<td>Full- or part-time</td>
</tr>
<tr>
<td>Disabled</td>
</tr>
<tr>
<td>Homemaker</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Disability compensation status, n (%)*</td>
</tr>
<tr>
<td>Straight leg raise or femoral tension, n (%)</td>
</tr>
<tr>
<td>Any neurologic deficit, n (%)</td>
</tr>
<tr>
<td>Reflexes—asymmetric depressed</td>
</tr>
<tr>
<td>Sensory—asymmetric decrease</td>
</tr>
<tr>
<td>Motor—asymmetric weakness</td>
</tr>
<tr>
<td>Perceive problem is getting worse</td>
</tr>
<tr>
<td>Definite surgical treatment preference, n (%)</td>
</tr>
</tbody>
</table>

*Receiving workers’ compensation, social security compensation or other compensation, or application for compensation pending.

2064 www.spinejournal.com

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
nonoperatively treated tended to have ongoing costs in years 3 and 4 that were higher than those observed for surgically treated patients, with this effect being most pronounced for individuals with DS, mainly because of productivity losses. By contrast, SPS patients had costs that were fairly comparable between surgically and nonoperatively treated patients, with costs during years 3 and 4 being slightly higher in operatively treated patients.

Net costs over 2 years were higher for surgically versus nonoperatively treated patients with reported differences, inclusive of initial and repeat surgery costs, of approximately $13,000 for SPS, $22,000 for DS, and $7000 for IDH under Medicare costing.\(^{11,12}\) Both DS and IDH patients who were nonoperatively treated tended to have ongoing costs in years 3 and 4 that were higher than those observed for surgically treated patients, with this effect being most pronounced for individuals with DS, mainly because of productivity losses. By contrast, SPS patients had costs that were fairly comparable between surgically and nonoperatively treated patients, with costs during years 3 and 4 being slightly higher in operatively treated patients.

Figure 1. Adjusted mean EuroQol-5D health state values and 95% confidence intervals over time by treatment received for (A) spinal stenosis, (B) degenerative spondylolisthesis with stenosis, and (C) intervertebral disc herniation disease groups. Treatment groups are compared assuming a common baseline value.

Figure 2. Mean costs and 95% confidence intervals by time period and treatment received for each disease group and type of cost: (A) direct medical costs and (B) indirect costs.

Copyright © 2011 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
Following effectiveness and cost patterns over time resulted in improved estimates of surgery's value—particularly for DS and IDH patients. These findings highlight the importance of following health and economic outcomes longitudinally to determine value over an extended time horizon.

Although the randomized clinical trial remains a cornerstone for comparative effectiveness research, it is widely recognized that alternative study designs, including observational cohorts, are necessary to support comparative evidence development for many diseases and population subgroups. Our analyses included SPORT’s randomized and observational cohorts and adjusted for factors known to affect treatment received because of the crossover between treatment groups. We acknowledge, however, that our analytic approach cannot control for any unmeasured differences between the two patient groups. To assess the potential impact of treatment selection on cost-effectiveness results in sensitivity analyses, we reported results for the observational and randomized cohorts separately and also undertook an analysis where surgeries occurring beyond 2 years were
removed from consideration. Mean costs per QALY gained remained fairly stable and in all cases fell within the 95% CI reported for the primary analysis.

Previous randomized and nonrandomized observational studies have shown a diminution in effect of surgery over time and cost-effectiveness reports have been based on decision-analytic models and/or incomplete longitudinal data. For example, a model-based analysis compared decision-analytic models and/or incomplete longitudinal time and cost-effectiveness reports have been based on studies have shown a diminution in effect of surgery over time. These data provide a basis for promoting fully informed clinical effectiveness and cost-effectiveness of surgery over time. These data provide a basis for promoting fully informed decision making between types of surgery involved decompression without fusion whereas 91% of surgeries in those with DS involved fusion. Likewise, our study was not powered to examine surgery by fusion type (instrumentation vs. not; type of instrumentation, etc.), yet the SPORT study represents the best available evidence to date with outcomes reported over 4 years. Finally, it is also important to emphasize that our results address the value of spine surgery in individuals with well-defined indications for surgery and cannot be generalized to other populations such as individuals with degenerative disc disease in whom surgery has become increasingly common.

Early cost-effectiveness results from SPORT suggested good value for surgery relative to nonoperative care for IDH and SPS, although the value of surgery for DS was not quite as favorable. However, it was noted that longer-term follow-up would be essential to fully characterize the cost-effectiveness of surgery for these specific indications. With follow-up over 2 additional years, it is evident that surgery for IDH has very favorable value, regardless of the approach to costing that is undertaken. The cost-effectiveness of surgery for stenosis improved slightly whereas the cost-effectiveness of surgery for DS improved markedly and now falls within the range of many commonly accepted medical and surgical practices. Continued follow-up of surgically and nonoperatively treated patients is necessary to provide further evidence regarding the clinical effectiveness and cost-effectiveness of surgery over time. These data provide a basis for promoting fully informed choice for patients with disc herniation or SPS with or without DS who face the difficult decision of whether or not to undergo spine surgery.

### TABLE 3. Sensitivity Analysis Results Shown as Mean Cost Per QALY Gained (95% Confidence Interval) for Surgery Relative to Nonoperative Care by Disease Group as Analytic Assumptions Are Varied

<table>
<thead>
<tr>
<th>Analytic Assumptions</th>
<th>Spinal Stenosis</th>
<th>Degenerative Spondylolisthesis</th>
<th>Intervertebral Disc Herniation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary analysis*</td>
<td>$59,400 ($37,059, $125,162)</td>
<td>$64,300 ($32,864, $83,117)</td>
<td>$20,600 ($4,539, $33,088)</td>
</tr>
<tr>
<td>Observational cohort</td>
<td>$48,300 ($12,011, $110,277)</td>
<td>$88,000 ($33,577, $151,986)</td>
<td>$28,200 ($5,283, $54,337)</td>
</tr>
<tr>
<td>Randomized cohort</td>
<td>$124,700 ($35,592, $459,143)</td>
<td>$51,600 ($18,885, $79,327)</td>
<td>$17,700 (save, $33,331)</td>
</tr>
<tr>
<td>Direct medical costs†</td>
<td>$52,100 ($29,504, $84,516)</td>
<td>$73,200 ($48,906, $87,650)</td>
<td>$26,500 ($16,784, $33,574)</td>
</tr>
<tr>
<td>Direct medical and worker productivity costs</td>
<td>$53,900 ($31,362, $91,012)</td>
<td>$72,200 ($48,124, $88,020)</td>
<td>$26,800 ($14,110, $33,863)</td>
</tr>
<tr>
<td>Higher surgery payment amount</td>
<td>$102,400 ($32,443, $115,557)</td>
<td>$122,500 ($33,199, $82,307)</td>
<td>$43,800 ($22,291, $57,563)</td>
</tr>
<tr>
<td>With mortality adjustment</td>
<td>$63,600 ($28,977, $96,795)</td>
<td>$55,100 ($29,220, $71,046)</td>
<td>$18,500 ($3,457, $29,514)</td>
</tr>
<tr>
<td>Initial surgery occurred within 2 years</td>
<td>$52,300 ($28,977, $96,795)</td>
<td>$55,100 ($29,220, $71,046)</td>
<td>$18,500 ($3,457, $29,514)</td>
</tr>
<tr>
<td>QALY estimation with SF-6D</td>
<td>$66,700 ($38,138, $122,820)</td>
<td>$79,300 ($43,314, $109,123)</td>
<td>$28,000 ($5,836, $46,913)</td>
</tr>
</tbody>
</table>

*Combined randomized and observational cohorts, all costs, Medicare surgery costs, no mortality adjustment, and EuroQol-5D scores used to estimate QALYs.
†Consistent with reference-case analysis recommended by panel on cost-effectiveness in health and medicine.

CI indicates confidence interval; QALY, cost per quality-adjusted life year.
Cost-Effectiveness of Spine Surgery • Tosteson et al

Key Points

- The cost-effectiveness of spine surgery for patients in three well-defined clinical groups (SPS, DS, and IDH) was assessed over 4 years among SPORT participants.
- For each group, the cost per QALY gained for surgery relative to nonoperative care improved compared with previously reported 2-year outcomes.
- Changes in the value of surgery when viewed over the longer time horizon were due to both durable QALY differences between surgically and nonoperatively treated patients and patterns of ongoing care costs.
- Our findings highlight the importance of including a contemporary comparison group in cost-effectiveness studies of spine surgery.

Acknowledgments

The authors thank Catherine C. Lindsay, SM, for her extensive work in the development of the cost weights used in this analysis, Tamara S. Morgan for creation of the patient diaries and for her assistance in preparing this manuscript, and Loretta Pearson for editorial assistance.

The authors thank the National Institute of Arthritis and Musculoskeletal and Skin Diseases (U01-AR45444), and the Office of Research on Women’s Health, the National Institutes of Health, and the National Institute of Occupational Safety and Health, the Centers for Disease Control and Prevention, for the funding. The Multidisciplinary Clinical Research Center in Musculoskeletal Diseases is funded by NIAMS (P60-AR048094).

References