Comparative effectiveness research in spine surgery

Kalil G. Abdullah, B.S.,1–3 Edward C. Benzel, M.D.,1–3 and Thomas E. Mroz, M.D.1–4

1Cleveland Clinic Lerner College of Medicine, 2Cleveland Clinic Center for Spine Health, and Departments of 3Neurological Surgery and 4Orthopaedic Surgery, Cleveland Clinic, Cleveland, Ohio

Comparative effectiveness research (CER) has impending significance for the field of spine surgery. This article outlines the rationale for comparative effectiveness research and reviews recommended priorities of spinal surgery emphasis. It also examines recent key studies of CER in the spine surgery literature and associated cost-effectiveness studies. It concludes with a discussion of the direction of CER in the spine surgery community.

(http://thejns.org/doi/abs/10.3171/2012.4.FOCUS1290)

**Key Words** • comparative effectiveness • cost-effectiveness • spine surgery

Comparative effectiveness research has risen to national attention for several reasons, principal among them being the strength and inertia of recently enacted legislation. 16,21 This legislation lays the foundation for both a governing body to monitor research effort and the provision of funding to individual investigators, so that this mechanism of research can be used and optimized for patient benefit. Comparative effectiveness research attempts to identify evidence that compares harm and benefit of different diagnostic, monitoring, and treatment options to improve delivery of care. The rationale for CER is to assist consumers and clinicians in making informed decisions that improve health care for individual patients as well as for populations of patients.

These concepts are not novel. Rather, they incorporate existing paradigms and attempt to insert economic analysis into clinical management strategies. Examples of the former include the Spine Patient Outcomes Research Trial (SPORT)31–33 and the vertebroplasty trials of 2009.7,19 In both cases, these randomized, controlled trials aimed to answer distinct questions regarding treatment efficacy. In SPORT, the investigators compared surgical versus conservative management for common spinal disorders, and the randomized, controlled trials published in 2009 examined the benefits of vertebroplasty versus a sham procedure in the treatment of osteoporotic fractures. These were notable trials because 1) Class I evidence had been lacking in the field of spine surgery and 2) the authors applied a prospective study design to answer important and common clinical questions. However, none of these studies was an example of CER. The authors analyzed efficacy of treatments, which is done under the controlled conditions of a clinical trial. But as this article will discuss, the SPORT investigators later applied economic and population-based concepts that show the potential of CER to expand on evidence gleaned from these types of clinical studies. That is, the effectiveness of the study was further examined, and this could only be accomplished by recording metrics such as the health-related quality of life during longitudinal routine care. This ability to understand both the efficacy and the effectiveness of spine surgery is a requisite of CER.

Importantly, the use of CER in spine surgery has already been examined, accepted, and supported independently by several medical governance consortia, including the Institute of Medicine (IOM), the Federal Coordinating

---

**Abbreviations used in this paper:**

- ACDF = anterior cervical discectomy and fusion
- CER = comparative effectiveness research
- CSM = cervical spondylotic myelopathy
- ICER = incremental cost-effectiveness ratio
- MIS = minimally invasive surgery
- mJOA = modified Japanese Orthopaedic Association
- ODI = Oswestry Disability Index
- QALY = quality-adjusted life year
- SF-36 = 36-item Short-Form Health Survey
- SPORT = Spine Patient Outcomes Research Trial
- TLIF = transforaminal interbody fusion
The investigators compared 4 years of data for patients who had been treated surgically and nonsurgically for spinal stenosis, degenerative spondylolisthesis, and disc herniation. As for their primary end point, they used the cost per QALY gained when comparing surgical to nonsurgical intervention.

The results of the study found that estimates of surgical value, relative to QALYs provided, continued to increase at 4 years after the primary event. This was particularly true for those patients with degenerative spondylolisthesis and disc herniation. The overall increases in QALY for these conditions over 4 years were 0.175 for spinal stenosis, 0.36 for degenerative spondylolisthesis, and 0.34 for disc herniation. The authors pointed out that while the effect was particularly pronounced at analysis 2 years after surgery, the incremental differences between 2 and 4 years were more modest (0.13 QALY for spondylolisthesis and herniation and only 0.05 for stenosis). This apparent lack of durable improvement highlights the importance of long-term, longitudinal analysis of data. The SPORT cost analysis at 2 years found considerable cost-effectiveness for disc herniation and spinal stenosis, but the evidence for degenerative spondylolisthesis was not as favorable. As mentioned above, the assessment at the 2–4 year period established the evolving outcomes and changes in QALY that affected interpretation of 2 of 3 of these diagnostic conditions.

A separate analysis published recently examined a subgroup of the SPORT patients with disc herniation in an attempt to examine the comparative effectiveness of surgical intervention versus conservative treatment for an individual diagnosis. The investigators analyzed the treatment effect of surgery, which they defined as changes in the ODI for surgical versus nonsurgical intervention. Overall, patients treated surgically for disc herniation improved when compared with nonsurgically treated patients, per ODI scores. However, their results suggested that at an individual patient level, certain factors contributed to a diminished treatment effect. The individuals whose treatment effect diminished were smokers, those with longer duration of symptoms prior to surgery without perisurgical worsening, and those with joint problems. These findings are by themselves not surprising, and many similar themes have been found in the regression analyses of retrospective studies of spine fusion in the past. The difference lies in the method of analysis being linked to a quality adjusted for each variable, rather than the previous standards of primary outcome, which may have included radiographic fusion or pain scales.

**The CSM Trial: Ventral versus Dorsal Surgery for CSM**

In 2011, Ghogawala et al. published the results of a nonrandomized, prospective trial at 7 sites over 2 years to determine the clinical effectiveness and cost of ventral versus dorsal decompression and fusion for the treatment of CSM. Fifty patients (28 undergoing ventral procedures and 22 undergoing dorsal procedures) were recruited; the 1-year follow-up rate was 92%. Four outcome scales were used to assess outcome measures (mJOA scale and ODI) and general health (the Physical Component Summary of the SF-36 and the EQ-5D (EuroQol Group) at 3, 6, and 12 months postoperatively. A limited economic analysis was
Spine surgery CER

performed with a subset of 41 patients and included only costs associated with the surgical admission. One year after surgery there was no statistically significant difference between the ventral surgery group and the dorsal surgery group with respect to quality of life. The ventral group had significantly favorable scores on the Neck Pain Disability Index and mJOA at 1 year. The mean values for hospital costs for ventral and dorsal surgery were $19,245 and $29,465, respectively. The confidence intervals were substantial for these figures and there was nearly $30,000 in variability in overall costs between institutions, but the mean cost per surgical group differed significantly.

The study's findings were bolstered by its rigid enrollment criteria and multicenter design. However, it was unable to address several important considerations. The complication profile of each of these groups was not provided, nor was the revision rate of each surgery or associated rationale for differing lengths of stay. Economically, the in-hospital costs are a convenient measure of differentiating between the 2 procedures, but ultimately follow-up visits, repeat imaging as a result of symptomatology, and delayed complications and revision surgeries are important in the consideration of a surgical cost profile. The authors acknowledged these limitations in the study, which was intended as a viability analysis for a larger, more sophisticated study. The study makes a strong case for a larger cohort and longer follow-up time, and it stands as one of the few studies (possibly the only one) led by a primarily neurosurgical contingent.

Interestingly, the study authors later performed a cost-effectiveness analysis utilizing an estimation of outpatient costs derived from cost diaries given to a subset of patients. Personal interviews and retrospective chart reviews were conducted to further derive cost estimates for patients enrolled in the original trial. An incremental cost-effectiveness ratio (ICER) was estimated by subtracting the cost of ventral surgery from that of dorsal surgery, and dividing that by the differential QALY gained in each operation (ICER = costs/QALY). The authors then compared 2 different cost-calculation methods: the cost-to-charge ratio (CCR) and the Medicare-reimbursement method (sum of the Current Procedural Terminology [CPT] and diagnosis-related group code costs). In the initial CCR analysis, dorsal surgery was found to be significantly more expensive than ventral surgery, but utilizing the Medicare-reimbursement method resulted in no significant difference between the costs for dorsal and ventral surgery. This was estimated to be a result of billing systems, as dorsal procedures necessitated a higher number of levels and instrumentation, as well as increased blood loss and length of stay (some of which allowed for higher billing parameters). Outpatient costs were still found to be significantly greater for those undergoing dorsal surgery, due to increased need for diagnostic testing and physical therapy. The results of ICER analysis again were more favorable for ventral surgery than dorsal surgery. It should also be emphasized that while charges may differ greatly based on various billing parameters (including region, health system, and so forth) true costs are a more important measure. The article importantly highlighted the complicating factors that are associated with proper estimation of cost-effectiveness utilizing differing methodology.

Spine Surgery Investigations Relating to CER

There have been several studies over the past several years that, while not necessarily fulfilling all criteria for CER, have attempted to systematically quantify the effectiveness of various spine surgeries. In the past few years, McGirt and colleagues have been particularly prolific in their analyses of the cost-effectiveness and cost-utility benefit of specific surgical techniques, with each investigation providing considerable length of follow-up. Several of these studies are summarized below.

In 2011, Adogwa et al. evaluated the cost-effectiveness of TLIF for Grade I degenerative spondylolisthesis. Forty-five patients with both leg and back pain underwent TLIF at a single institution. The authors found that the total cost (direct and indirect) of TLIF was $36,836 and the adjusted cost of TLIF per QALY gained was $42,854. These results were consistent with values reported by Glassman et al. of $53,914 for 5-year postoperative direct and indirect costs in 80 patients. The study by Adogwa et al. had several limitations, the first of which was the lack of a nonoperative group that would have allowed for comparison. This was addressed by the authors as a study strength rather than a weakness, since SPORT “suffered” from the crossover between the medical and surgical groups. This ability of SPORT patients to cross over meant that patients who remained in the medical group had inflated utility measures due to their desire to avoid surgery, thus decreasing the perceived utility of surgery relative to the nonoperative group. Nonetheless, the ability to compare nonoperative courses and quality of life remains important in assessing effectiveness over time.

The study by Adogwa et al. succeeds very well in a comparative sense; that is, the authors point out the value of these procedures relative to other commonly accepted surgical practices in the orthopedic world. The investigators point out that a total knee arthroplasty may cost $59,262 per QALY, and total hip arthroplasty may reach up to $80,000. With the notion that cost per QALY gained is acceptable at below $100,000, the cost reported for these spinal interventions is well within a favorable range. In 2012, the group determined that the cost per QALY gained of laminectomy and extension of instrumented fusion for adjacent segment disease was $62,995. While this does not address whether the primary surgical intervention was appropriate nor sum the QALY of the primary and revision surgeries, it supports the importance of valuating spinal intervention proactively.

Also in 2011, the same group evaluated the cost-effectiveness of minimally invasive versus open TLIF in 30 patients with the diagnosis of degenerative spondylolisthesis, who were undergoing the procedure for back and leg pain. Quality-adjusted life years were calculated from the EQ-5D, and the differences in mean cost were calculated by ICER. At 2-year follow-up, the number of QALYs gained was similar in the 2 groups, and the MIS group had a median cost savings of $8,731, although this was not statistically significant. The authors then followed their initial study with an analysis of return to work and
duration of narcotic use, again at 2-year follow-up and in the same 30 patients.1 They found that return to work and duration of narcotic use were significantly shorter in the MIS group than in the open surgery group (8.5 days vs 17.1 days and 2 weeks vs 4 weeks, respectively).

When taken individually, these investigations by McGirt and colleagues do not qualify as CER in its strictest definition, but the overall information provided allows a window into both the efficacy and efficiency of individual surgical treatments based on QALY and self-reported health, with particular attention to long-term follow-up.

Several other authors have attempted to interpret the cost-effectiveness of various spinal surgeries. In 2005, Angevine et al.6 performed a retrospective review of 78 patients to determine cost per QALY of single-level ACDF. They found that ACDF with allograft and plating had a relative benefit when compared with ACDF with autograft at a cost per QALY of $32,560. Tumialán et al.7 examined the cost-effectiveness of posterior cervical foraminotomy versus ACDF in a military population by analyzing a variety of factors that included direct costs as well as return to work. They found that the indirect costs of surgery were approximately $10,000 higher in the ACDF group than in the posterior foraminotomy group. This difference was largely due to the fact that the average return to work time was 19 weeks longer in the ACDF group. Like the papers above, these studies themselves do not qualify as CER, as many do not incorporate more than one type of effectiveness and focus mostly on cost, but they represent important contributions in what is still a nascent field.

Tubular versus Conventional Microdiscectomy

In 2011, van den Akker and colleagues presented a cost-utility analysis and multicenter, double-blind, randomized, controlled trial of 325 patients undergoing discectomy for sciatic symptoms.8 They used the EQ-5D and the SF-6D (derived from the SF-36) taken at defined intervals between 2 and 52 weeks after randomization. They then estimated the costs based on cost diaries given to patients, with an overall compliance rate of 90% at 52 weeks. The authors were unable to find any significant differences in QALY, direct health care, or societal costs between the 2 procedures. This led them to conclude that it was unlikely that tubular discectomy was any more cost-effective than traditional microdiscectomy. The validity of this conclusion may be hampered by surgeon preference at any one of the 7 surgical sites. Learning curves are steeper in MIS procedures and if costs are not found to be significantly different at any level of analysis, researchers would focus their attention on the direct surgical costs, which in this study were based on duration of operating room use. Thus, more experienced surgeons could alter the direct cost component, which would introduce bias toward lower cost.

Comparative Effectiveness in Spine Surgery–Related Fields

There has been substantial movement toward CER in the fields of pain management, anesthesia, and interventional radiology. While these are beyond the scope of this manuscript, they are mentioned here briefly for the sake of completeness. These investigations have included multicenter, randomized, controlled trials of diagnostic facet blocks,9 treatment of chronic thoracic pain,10 and balloon kyphoplasty.11

Conclusions

Conceptualizing that the field of spine surgery could be significantly altered by CER may be difficult. There also may be objections to CER for its perceived inability to measure individual patient situations or because its reduction of decision making to discrete long-term effectiveness is disconcerting.12 These concerns are well-founded, but they are secondary to the reality of CER’s arrival as a measure of the efficiency and value that spine surgery practitioners may provide the patient (or, often in CER literature, the consumer). Although the recommendations derived from CER have been prohibited from utilization as a direct basis for alteration in health care funding policy in the US, this has not been the foreign CER experience. It is reasonable and practical that ultimately a governmental body whose sole purpose is to efficiently maximize resource utilization would look (either officially or unofficially) to research that exists to provide a basis of efficacy, cost-utility, and effectiveness over time to assist in their decision-making process. Comparative effectiveness research has become a governmental priority, and it is here to stay.

The spine surgery community is more likely to be affected by these developments than any other neurosurgical specialty, and possibly more so than many other surgical fields. Recent attention paid to the proper indications for spine surgery and the intimacy of relationships between spine surgeons and device manufacturers has resulted in public, industry, and governmental consideration of reimbursements and best-practice guidelines.3 These changes reflect a growing call for accountability to strong scientific process and solid, evidence-driven practice. Comparative effectiveness research provides opportunities for the spine surgery community to not only participate in the CER enterprise and address these issues, but also to become leaders in the discussion of what constitutes the proper downstream uses of CER. Comparative effectiveness research is being formed and molded mostly by nonsurgeons and certainly non–spine surgeons, but the implications of its reach certainly have repercussions for our field.17

The federal government has allocated $1.1 billion to conduct CER, including $400 million distributed directly to the National Institutes of Health, and opportunities to answer important questions in our field abound. Answering these questions requires a solid understanding of clinical spine surgery as well as the basic tenets of CER. Although CER may be foreign to many spine surgeons, the involvement of surgeons in CER is essential to the sustainability of spine surgery as a robust and innovative surgical field.

Disclosure

Dr. Mroz reports having a consultant relationship with Globus.
Spine surgery CER

Medical, Inc., owning stock in Pearl Diver, Inc., and receiving speaking honoraria from AO Spine.

Author contributions to the study and manuscript preparation include the following. Conception and design: Abdullah. Analysis and interpretation of data: Abdullah. Drafting the article: Abdullah. Critically revising the article: all authors. Reviewed submitted version of manuscript: Abdullah. Administrative/technical/material support: Mroz, Benzel. Study supervision: Mroz, Benzel.

References


