

SPORT: What Neurosurgeons Need to Know

Peter D. Angevine, M.D., M.P.H., and Paul C. McCormick, M.D., M.P.H.

“The results of the three parts of the Spine Patient Outcomes Research Trial (SPORT) were published within the 18 months.” In November 2006, the reports of the randomized, controlled trial (RCT) and observational cohort study of herniated lumbar disc (HLD) were published in the *Journal of the American Medical Association*.^{8,9} In May 2007, the *New England Journal of Medicine* published the results of the RCT of the surgical versus nonoperative treatment of degenerative spondylolisthesis.⁷ Finally, in February 2008, the results of the lumbar spinal stenosis study were published, also in the *New England Journal of Medicine*. Practicing neurosurgeons are likely to be asked about the results of these studies because they were covered extensively in the popular press, particularly the HLD RCT. Neurosurgeons should be able to discuss with their colleagues and patients the strengths and weaknesses of these studies to help them understand what information the studies provided and, perhaps more importantly, what information they did not.

Organized neurosurgery objected to the fundamental hypothesis and design of the SPORT trial in 2003.³ The concern among the neurosurgical leadership arose from the overly simplistic and irrelevant formulation of the study question. By its design, the SPORT HLD RCT would provide data to either reject or not reject the following null hypothesis: “there is no difference in outcome between patients with painful HLD who, after at least 6 weeks, receive continued nonoperative care and those who undergo lumbar microdiscectomy.” The implication of this statement, that nonoperative care and lumbar microdiscectomy are exchangeable interventions and that there exists one superior and one inferior treatment for a disorder as individual and subjective as painful HLD, represents a fundamental misunderstanding of the diagnosis and the patients. The study proceeded without the participation of neurosurgeons.

Herniated Lumbar Disc Randomized, Controlled Trial and Cohort Study

The results of the SPORT HLD study have been reviewed elsewhere and are only briefly summarized here.^{4,5}

Approximately 62% of patients (1244 of 1991 eligible) offered participation in the study enrolled; of those, 40% (501 of 1244) consented to randomization, raising serious questions about the generalizability of the study. This potential problem was anticipated in a letter to Dr. Weinstein, the principal investigator of SPORT, written by the leadership of neurosurgery: “patients with more acute, severe symptoms . . . are less likely to agree to randomization [than patients with less severe symptoms].”³ On average, the patients with severe symptoms that enrolled in SPORT declined randomization; most chose surgical treatment as part of the nonrandomized arm of the study. The overall mean Oswestry Disability Index for patients in the RCT and the observational cohort were 46.9 and 51.2, respectively ($P < 0.001$). Also, a greater proportion of patients in the observational cohort reported a perception that their problem was worsening compared with patients in the RCT (43% versus 34%, $P < 0.001$).

Patients who agreed to undergo randomization were randomized in blocks by study center only; no allowance was made for known predictors of outcome of surgery. The Maine Lumbar Spine Study (MLSS), for example, had convincingly demonstrated that patients with less severe symptoms tend to select nonoperative management.¹ Additionally, those patients with milder symptoms that do undergo surgery have less benefit compared with patients who have more severe symptoms and have an operation. In the MLSS, approximately 20% of the patients with “mild” symptoms (based on the Sciatica Bothersome Index) selected surgical treatment, whereas almost three-fourths of those with severe symptoms chose to undergo surgery. A comparison of the outcomes of the operative and nonoperative treatment of patients with mild and moderate symptom severity demonstrated that surgery was not associated with a significant difference in outcome among patients with lesser symptom severity. Patients with moderate symptom severity who underwent surgery, on the other hand, had statistically significant improvements in Sciatica Frequency Index, quality of life, and satisfaction compared with those who continued with nonoperative treatment. The response to surgery is not, therefore, uniform, but varies depending on the severity of the patient’s symptoms. Grouping all patients together for treatment and

analysis as the SPORT investigators did, particularly with an unbalanced group of severely and mildly symptomatic patients, ignores this fact and biases the study toward the null.

Intent-to-treat (ITT) analysis is a methodological cornerstone of randomized clinical trials. Only the as-randomized groups are, on average, balanced in the distribution of measured and unmeasured factors that may be associated with response to treatment. With high proportions of compliance, and therefore a high correlation between the as-assigned and as-treated groups, ITT analysis offers the strongest protection against bias of all the methods of statistical analysis. As compliance decreases, however, the inference that treatment allocation is a proxy for treatment received becomes less valid and the results of ITT analysis reflect the effect of treatment assignment not treatment delivered. Within the RCT, there was an extraordinarily high bidirectional crossover rate, effectively negating the random allocation of patients as a proxy for treatment received. At 2-year follow-up, 140 of the 232 patients in the group assigned to the surgical arm who had any follow-up data (60%) had undergone surgery. At the same time point, 107 of the 240 analyzed patients (45%) in the nonoperative treatment arm crossed over and underwent surgical treatment (Fig. 8.1).

With high proportions of crossovers in both treatment assignment groups, any difference in outcome as a result of the treatment itself will be obscured as the groups become more similar with respect to the experimental variable that is supposed to differ between them (Fig. 8.2). Another way of considering this problem is that crossovers effectively decrease the power of a study. That is, with increasing proportions of patients who cross over from the assigned treatment to the alternative intervention, the likelihood of a study detecting a difference in effectiveness of a given size between the treatments decreases. With the high crossover in the

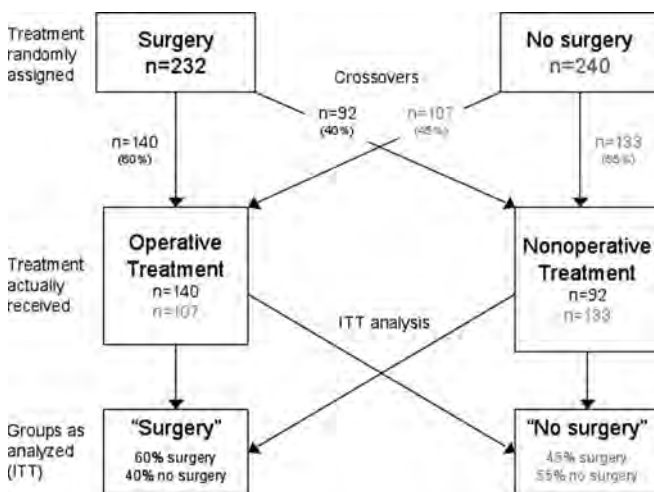


FIGURE 8.1. Schematic of crossovers in SPORT RCT.

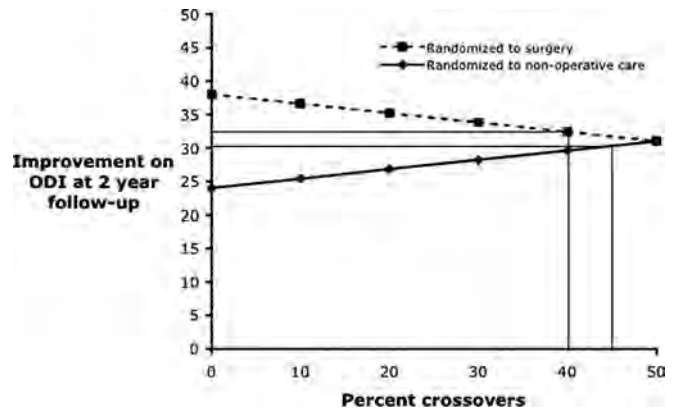


FIGURE 8.2. Illustration of the effects of crossover on effect size with ITT analysis. The effect sizes (difference in improvement in the Oswestry Disability Index [ODI]) of operative and nonoperative interventions for zero crossovers are based on the cohort study. With 50% bidirectional random crossover, the apparent effect size would be zero. The actual crossover proportions from the SPORT RCT are indicated, 40% in the group randomized to surgery and 45% in the group randomized to nonoperative treatment. This graphic result is very close to the actual result of an apparent effect size of 2.7 points on the ODI in the ITT analysis of the RCT.

SPORT RCT, there was an approximately 7% probability that a difference of 10 points or greater on the relevant SF-36 scales would be found if such a difference existed (Fig. 8.3).

The only valid conclusion that can be drawn from the ITT analysis is that of patients who agreed to enroll in the RCT, assignment to receive surgical or nonsurgical treatment for symptomatic HLD is not associated with clinically or statistically significant differences in outcome at 1 year. Unfortunately, most expert and lay commentators on the

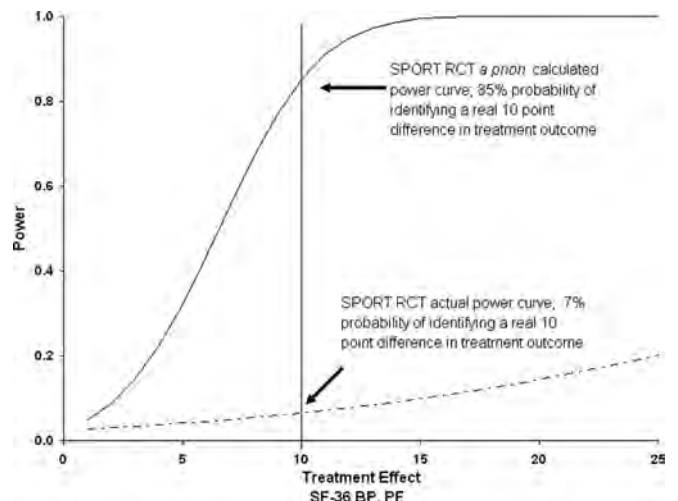


FIGURE 8.3. Effect of crossovers on power of SPORT HLD RCT.

study missed this critical point and published articles implying that patients who received surgery or nonoperative care fared about the same; therefore, surgery was unnecessary if one could “wait it out.”

Most commentators failed to note that the RCT published report did mention that as-treated analysis demonstrated that patients who underwent surgery had significantly improved clinical outcomes at all time points compared with those who continued to receive nonoperative care. Nor was much attention paid to the observational cohort study that directly followed the RCT in the *JAMA* issue, where an even clearer picture emerged. Despite having, on average, greater pain and disability at baseline compared with the nonoperative cohort, patients who underwent lumbar microdiscectomy had statistically and clinically significantly better outcomes. Observational cohort studies are, of course, potentially subject to greater allocation bias, the systematic difference in outcome resulting from differences in treatment assignment than randomized trials.

A close examination of the patients’ baseline parameters in the RCT and observational study reveals an explanation of the crossover population. Although the patients in the RCT in the surgery and nonoperative arms had comparable baseline pain and disability scores that were intermediate between the surgery and nonsurgery observational cohorts, segregating the RCT patients by treatment received revealed a different picture. The mean baseline pain and disability scores and the proportion of patients who felt that their condition was worsening among patients who underwent surgery in the RCT (from either assigned group) and the observational study were strikingly similar (*Table 8.1*). A similar finding is noted for the patients who underwent nonoperative treatment. Consistent with logical expectations, patients who had greater pain or disability or who thought their condition was worsening tended to select surgery,

whereas those who were less severely symptomatic or who were improving opted to continue with nonoperative management, regardless of treatment assignment.

What lessons, then, did we learn from SPORT? We learned that patients in the U.S. private healthcare system with symptomatic HLD will not comply with randomization to surgical or nonsurgical treatment in a sufficiently high proportion to allow the implementation of a valid RCT comparing operative and nonoperative treatment. We learned that, on average, patients with greater pain or more compromised function or who think they are getting worse more often select surgery for treatment of their painful HLD than those with less pain or disability or who think they are improving. “Neither of these were particularly surprising results.”

Unfortunately, there are other potentially important lessons that we did not learn. We did not learn if the neurological outcome of a patient with a stable neurological deficit such as a foot drop is better with lumbar microdiscectomy or not. We did not learn any new factors that are associated with a higher likelihood of good (or poor) outcome with surgery or with nonoperative treatment. Finally, we did not learn, despite what one commentator stated, that concerns about the development or progression of neurological deficits with nonoperative treatment were “simply not borne out.”³ That conclusion could only have been validly reached had every patient remained in his or her randomly assigned treatment group for the entire study. Instead, we learned that long-established treatment principles are effective; patients may be (and are) managed nonoperatively with little risk and a reasonable expectation of improvement if they may elect to undergo surgery at any time. This key distinction is at variance with “sound bite” summaries of the study that incorrectly attributed outcomes to assigned, rather than received, treatment.⁴

TABLE 8.1. Comparison of baseline pain and disability scores of patients in the RCT and observational study segregated by treatment received

	RCT: Assigned to Surgery		RCT: Assigned to Nonoperative Management		Observational Study	
	Surgery (n = 140)	No Surgery (n = 92)	Surgery (n = 107)	No Surgery (n = 133)	Surgery	No Surgery
ODI ^a	51.7 (20.9)	41.1 (20.7)	52.1 (19.2)	41.6 (20.6)	56.7 (18.9)	35.9 (20.1)
SF-36 ^b						
<i>Bodily pain</i>	24.1 (16.7)	31.7 (20.2)	24.1 (16.8)	28.9 (17.7)	21.2 (15.8)	36.2 (20.3)
<i>Physical functioning</i>	35.9 (24.0)	45.6 (25.3)	33 (22.9)	44.1 (26.9)	30.8 (23.0)	52.5 (25.9)
<i>Worsening health trend^c</i>	58 (41)	24 (26)	44 (41)	35 (26)	272 (52)	39 (30)

^aLower score indicates less severe symptoms; range 0–100; numbers are mean (SD).

^bHigher score indicates less severe symptoms; range 0–100; numbers are mean (SD).

^cNumbers given as number (percent).

RCT, randomized, controlled trial; ODI, Oswestry Disability Index; SD, standard deviation.

Degenerative Spondylolisthesis Randomized, Controlled Trial

In May 2007, the second part of SPORT, an RCT of the surgical versus nonoperative treatment of lumbar degenerative spondylolisthesis, was published. Patients with at least 12 weeks of persistent symptoms (neurogenic claudication or lumbar radicular pain) and radiographically confirmed lumbar stenosis and degenerative spondylolisthesis were eligible for enrollment. Six hundred and seven of the 892 eligible patients enrolled in the study; 304 agreed to randomization. Similar to the HLD RCT, the spondylolisthesis trial was plagued by a high level of bidirectional crossover. At 2 years, 64% of patients assigned to surgery and 49% of those assigned to nonoperative treatment had undergone surgery. Patients who underwent surgery, whether in the RCT or the observational cohort, tended to have worse physical function, greater disability, and more pain compared with patients who continued with nonoperative treatment during the study period.

The median percent of slip was 15%; the range was 1 to 37%. Approximately 95% of surgical procedures were decompressions with arthrodesis. Just over one-fourth of the fusions were noninstrumented. Almost 75% of all operations, therefore, were decompressions with instrumented fusions.

Because of the high proportions of crossover in both directions, the ITT analysis was biased toward the null and no statistically significant difference in outcome between the randomized groups was found. As-treated analysis was performed combining all patients (in the RCT and observational arms) who underwent each treatment. This analysis demonstrated a statistically and clinically significant benefit for surgery for all primary and secondary outcomes and at all assessment time points.

Although, as mentioned previously, as-treated analysis is potentially subject to bias, in the presence of high bidirectional crossover, it is likely to produce a more accurate estimate of the effect of the intervention itself. With this caveat, this study provides evidence that for patients similar to those included in this study, surgery offers a benefit in clinical outcome at up to 2 years compared with nonoperative treatment.

CONCLUSION

Despite the prediction of Steven N. Katz, the director of the National Institute of Arthritis and Musculoskeletal and Skin Diseases in 1999, that SPORT would “for the first time, [provide] scientific evidence regarding the relative effectiveness of surgical versus nonsurgical treatment of these commonly diagnosed lumbar spine conditions,” this multimillion

dollar study failed to do so.⁶ The study was limited from the outset by simplistic, “winner take all” hypothesis and the likelihood of a nonrepresentative patient sample. Exacerbating these problems was the authors’ emphasis on the ITT analysis, which was severely compromised by the large number of patients that crossed over in both directions.

We learned from SPORT mainly what we already knew. Patients with relatively mild symptoms from lumbar HLD and those who are improving over time will tend to choose to continue with nonoperative care and will, in general, do well at long-term (1 to 2 years) follow-up. Those with more severe disease and those who are worsening despite nonoperative care will have a tendency to decide to undergo surgery. These patients will, on average, obtain a significant, lasting benefit from the procedure. Patients’ choices regarding their treatment are unlikely to be significantly altered by random allocation in a randomized clinical trial; therefore, the useful information obtained from such a study is minimal.

Disclosure

The authors did not receive financial support in conjunction with the generation of this article. The authors have no personal or institutional financial interest in the drugs, materials, or devices described in this article.

REFERENCES

1. Atlas SJ, Deyo RA, Keller RB, Chapin AM, Patrick DL, Long JM, Singer DE: The Maine Lumbar Spine Study, part II: 1-year outcomes of surgical and nonsurgical management of sciatica. *Spine* 21:1777–1786, 1996.
2. Carragee E: Surgical treatment of lumbar disk disorders. *JAMA* 296:2485–2487, 2006.
3. Dunsker SB, Awad IA, McCormick PC: Spine patient outcomes research trial. *J Neurosurg* 98:1150–1152, 2003.
4. Kolata G: Study questions need to operate on disk injuries. Available at <http://www.nytimes.com/2006/11/22/health/22spine.html>. Accessed 11/22/2006.
5. McCormick PC: The Spine Patient Outcomes Research Trial results for lumbar disc herniation: A critical review. *J Neurosurg Spine* 6:513–520, 2007.
6. National Institutes of Health: NIAMS funds multicenter study of surgical vs. nonsurgical treatment of back pain. Available at: http://www.niams.nih.gov/News_and_Events/Press_Releases/1999/09_20.asp. Accessed July 3, 2008.
7. Weinstein JN, Lurie JD, Tosteson TD, Hanscom B, Tosteson ANA, Blood EA, Birkmeyer NJO, Hilibrand AS, Herkowitz H, Cammisia FP, Albert TJ, Emery SE, Lenke LG, Abdu WA, Longley M, Errico TJ, Hu SS: Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med* 356:2257–2270, 2007.
8. Weinstein JN, Lurie JD, Tosteson TD, Skinner JS, Hanscom B, Tosteson ANA, Herkowitz HN, Fischgrund JS, Cammisia FP, Albert TJ, Deyo RA: Surgical vs nonoperative treatment for lumbar disk herniation: The Spine Patient Outcomes Research Trial (SPORT) observational cohort. *JAMA* 296:2451–2459, 2006.
9. Weinstein JN, Tosteson TD, Lurie JD, Tosteson ANA, Hanscom B, Skinner JS, Abdu WA, Hilibrand AS, Boden SD, Deyo RA: Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): A randomized trial. *JAMA* 296:2441–2450, 2006.