The Spine Patient Outcomes Research Trial results for lumbar disc herniation: a critical review

PAUL C. MCCORMICK, M.D., M.P.H.

Department of Neurosurgery, Columbia University College of Physicians and Surgeons, New York, New York

The long-anticipated results of the Spine Patient Outcomes Research Trial (SPORT) were recently published in the Journal of the American Medical Association. In this trial the investigators compared operative and nonsurgical care in patients with symptomatic lumbar disc herniation. Despite the expenditure of several million dollars on this multicenter, prospective, randomized, controlled clinical trial, the SPORT investigators admitted, “conclusions about the superiority or equivalence of the treatments under study are not warranted based on the intent-to-treat analysis.” In the present article the author provides a critical review of the SPORT formulation and hypothesis, study design and methodology, and results and interpretations in an attempt to explain why the authors of this study were unable to assess the study’s only intended null hypothesis that there would be no difference in outcomes between operative and nonsurgical management of herniated lumbar discs. Issues related to misrepresentation and misinterpretation of the SPORT results for herniated lumbar discs are also assessed.

KEY WORDS • Spine Patient Outcomes Research Trial • herniated lumbar disc • surgery • nonoperative treatment • randomized controlled trial

On September 19, 1999, the NIAMS announced federal funding of the multicenter SPORT in which the intention was to assess the relative efficacy of surgery compared with nonsurgical care in the treatment of patients with lumbar disc herniation, lumbar spinal stenosis, and degenerative lumbar spondylolisthesis. In a for-immediate-release press statement, Dr. Steven Katz, Director of NIAMS, stated unequivocally that, “Based on this trial, we shall, for the first time, have scientific evidence regarding the relative effectiveness of surgery versus nonsurgical treatment of these commonly diagnosed lumbar spine conditions.” Now, more than 7 years later, following the enrollment of 1244 patients and the expenditure of $13.5 million, the initial results of the lumbar disc herniation study have been published in the JAMA. In a disappointing declaration, the SPORT investigators admitted, “conclusions about the superiority or equivalence of the treatments under study are not warranted based on the intent-to-treat analysis.”

How could this have happened? How could the authors of a large, multicenter, National Institutes of Health–reviewed and –funded multimillion-dollar study, which used the gold-standard RCT study design, fail to validly assess even the single intended a priori null hypothesis of no difference between surgical and nonsurgical treatment for herniated lumbar discs? The answer is simple. Weaknesses in the study hypothesis, lack of a priori stratification of known prognostic variables, failure to account for the anticipated high rate of treatment crossovers in the study design, flaws in the data analysis, and nonrepresentativeness of the SPORT RCT population virtually guaranteed the inability of this study to validly test its only intended hypothesis. The authors of numerous journal publications have documented these concerns, none of which was adequate-
ly addressed or even referenced by the SPORT investigators. Thus, this disappointing result was both predictable and avoidable.

More troubling than the poorly formulated SPORT hypothesis and the biased-toward-the-null study design is the inaccuracy in the interpretation of the SPORT results by the lay press, the medical community, and even peer-reviewed sources, the latter two of which hold the public trust for scientific and intellectual integrity. For example, the JAMA, in their prepublication press release entitled “Rundominated study indicates that patients with herniated disc improved with or without surgery,”89 claimed in the opening paragraph, “Patients with lumbar disk herniation who had surgery or nonoperative treatments showed similar levels of improvement in the reduction of pain over a 2-year period, according to a randomized trial in the November 22/29 issue of JAMA.” The NIAMS website, reporting on the SPORT RCT under the headline, “Study shows patients with herniated disks improve over time— even without surgery,”19 claimed: “(SPORT) researchers found that those who forewent surgery for nonoperative care fared similarly to those who had surgery.” The NIAMS Director Steven I. Katz, M.D., Ph.D., is quoted as saying, “nonoperative therapies may offer benefits that are similar to if not equal for patients who cannot or elect not to have surgery.” On ABC News,9 under the headline “Study shows that patients get same benefits regardless of whether or not they choose the knife,” Dr. Richard Deyo, a coprincipal investigator of SPORT, was quoted as saying, “This trial confirms some of the earlier findings, suggesting results are similar with modern surgical and non-surgical treatment techniques.” In this report Dr. William Richardson, an orthopedic spine surgeon from Duke University, was also quoted as follows: “This confirms what we already know, patients get better quicker with surgery, but the long term results are no different.” The December 19th edition of Neurology Today reported, “Patients who underwent surgery achieved slightly better results more quickly. However, the differences were not statistically significant...”19

Surprisingly, these statements implying comparisons between actual treatments received were made despite the clear admonition by the SPORT investigators that such direct treatment comparisons are not valid (for example, “conclusions about the superiority or equivalence of the treatments under study are not warranted based on the intent-to-treat analysis”).23 None of these statements accurately characterizes the findings of the SPORT RCT because each erroneously interprets the ambiguous term “treatment group” used in the SPORT RCT publication as synonymous with the actual treatment received. It was not. Due to the planned intent-to-treat analysis and the remarkably high crossover rate, for example, 45% of the patients in the nonoperative “treatment group” actually underwent surgery, and the surgical results were credited to nonoperative management.

Given the unclear language in the SPORT RCT publication and the inaccurate characterization of the results in the medical community, it should come as no surprise that much of the lay press erroneously reported essential equivalence of outcome irrespective of treatment.5,7,8,11,13,16,17,20 Newsweek,20 for example, reported that “researchers said the difference in outcomes between the two approaches (actual surgery vs. actual nonoperative treatment) were small and not statistically significant.” The New York Times10 claimed: “Study Questions Need to Operate on Disk Injuries,” in their front page headline on the JAMA SPORT publication. National Public Radio17 claimed: “nonoperative remedies are almost as effective as surgery in relieving pain,” and MSNBC23 reported: “two big government studies [that is, SPORT] on back surgery for painful herniated disks show no clear-cut reason to choose an operation over another treatment.” These statements are inaccurate because they also assume, incorrectly, that the intent-to-treat analysis directly compared the actual treatments received.

In the remainder of this editorial I will attempt to critically analyze the limitations and the flaws of the two SPORT studies on herniated lumbar discs that have resulted in a poorly formulated, biased-toward-the-null, and misinterpreted study. In the end, it is unfortunate that so little valid and relevant inferential data resulted from what was obviously an extraordinary effort by the SPORT investigators.

The SPORT Study Summary

Due to the prohibitively high crossover rate in the SPORT RCT, Weinstein and colleagues23 were unable to validly test the study’s only stated a priori null hypothesis that there was no difference between operative and nonoperative treatment of lumbar disc herniation. Instead, only the modified hypothesis of whether an initial management strategy of offering nonoperative treatment produced different results than an initial management strategy of offering surgery for patients with painful lumbar disc herniation could be validly assessed in the SPORT. Not surprisingly, the answer to this clinically unimportant question was no. It is equally important to appreciate what the SPORT study did not conclude, despite media reports to the contrary. Specifically, the authors of the study did not establish equivalence between operative and nonoperative treatment for lumbar disc herniation, nor did they demonstrate that every patient with a symptomatic lumbar disc herniation is equally likely to experience the same outcome eventually, irrespective of the treatment received. In both the SPORT observational cohort and in the as-treated RCT analysis,22,23 for example, patients who underwent surgery had not only more rapid symptom relief but experienced both clinically important and statistically significant superior outcomes than patients treated nonoperatively on virtually all primary and secondary outcome measures; these differences persisted at every follow-up interval throughout the entire 2-year study period. Thus, although one can conclude that the initial treatment offered had no differential effect on clinical outcomes, one cannot conclude, due to the high crossover rate, that the actual treatment received also produced equivalent outcomes.

Even the claims that patients in both as-treated groups experienced “substantial improvement” by 2 years in both SPORT publications,22,23 albeit true, are nevertheless misleading in light of the clinically important and statistically significant superior outcomes achieved by the as-treated surgical patients in both the SPORT RCT and observational cohort. It is probably safe to assume that most patients would prefer, or at least be given the choice of, maximum
improvement over substantial improvement, particularly when the difference between the two outcomes is of a clinically important magnitude. As Dr. Weinstein clearly states, the goal of SPORT was simply to “find the truth for patients.” Additional concerns regarding the SPORT formulation, design, analysis, and interpretation appear in subsequent sections.

Study Hypothesis

Formulation of a valid and relevant study hypothesis is essential for prospective investigational studies, particularly expensive and resource-intensive multicenter randomized trials. Only the a priori hypothesis can be validly assessed by the investigators. The hypothesis must take into account the nature of the condition and treatments under study, previously accumulated empirical experience, and prior observational and experimental investigation.

The purpose of the SPORT RCT was simple: to test the hypothesis that there was no difference between operative and nonoperative treatment in patients with herniated lumbar discs. The SPORT investigators tested a dichotomous hypothesis in which they sought a categorical yes or no answer to the question of whether surgery is better than nonoperative treatment for lumbar disc herniation. This finding could then, presumably, be extrapolated to all patients with the condition. This hypothesis, however, is problematic on several levels. First, based on the nature of lumbar disc herniation and the treatments under study, the intended SPORT hypothesis seems neither valid nor relevant because such an elementary hypothesis assumes, incorrectly, that lumbar disc herniation is a homogeneous condition, that surgery and nonoperative therapy are competitive treatments, and that, in the absence of bias and confounding elements afforded by the randomized trial design, there is a uniform response to each treatment. Second, what would be the possible policy implications if surgery, on average, produced superior outcomes, as it did both in the SPORT observational cohort and the as-treated RCT analysis? Would all patients with herniated lumbar discs be offered surgery at their initial presentation? Of course not. Alternatively, had nonoperative treatment, on average, been proven superior, or even equivalent, would surgery no longer be offered as a treatment for a painful herniated lumbar disc?

Such are the only choices that the simplistic winner-takes-all a priori hypothesis of the SPORT RCT allows.

Neither the intended nor the modified SPORT hypotheses (that is, which is the better initial treatment strategy for a painful herniated lumbar disc?) are particularly relevant in contemporary practice because surgical and nonsurgical care for painful lumbar disc herniation are not parallel, competing, or interchangeable treatments, nor is symptomatic lumbar disc herniation a homogeneous disease with respect to pain, neurological dysfunction, disc location and morphological features, clinical course, natural history, treatment response, or patient preference. Relief of pain and restoration of function with as little risk as possible is the fundamental management strategy that nearly all physicians apply to cases of painful lumbar disc herniation. With few exceptions, therefore, all patients presenting with a painful herniated lumbar disc are initially treated with nonoperative therapy. Many, if not most, patients will improve with this conservative treatment. Patients will rarely choose, and physicians will usually not offer, a surgical option to patients if their pain is of short duration, of mild to moderate intensity, or in the process of improving. Indeed, many of the patients who agreed to enroll in the SPORT RCT, particularly those with mild and/or improving symptoms, would likely not have been offered and/or have elected to undergo surgical treatment outside of the SPORT trial. Those patients in whom symptoms either do not improve or worsen, despite a reasonable course of nonoperative treatment, may choose the option of surgery. A treatment strategy that poses surgical and nonsurgical therapies as complementary treatments and takes into account the heterogeneous clinical aspects of lumbar disc herniation and the benign, self-limited nature for most symptomatic herniated lumbar discs, represents the current best practice for painful lumbar disc herniation and is supported by virtually all published evidence-based clinical guidelines for low-back pain. Indeed, even the evidence-based videotape shown to each patient in the SPORT observational cohort and RCT provided this information, including an emphasis on the often self-limited nature of lumbar disc herniation and the ultimate equivalent outcomes for operative and nonoperative management. The reality is that in contemporary clinical practice, due to the heterogeneity of the condition under study and the different circumstances under which each treatment is applied or chosen, there is, appropriately and by design, an intentional bias in the application of nonoperative and operative treatments for lumbar disc herniation. To eliminate this rational patient treatment selection bias in an ill-conceived dichotomous SPORT randomization scheme reflects a fundamental flaw in the SPORT study formulation.

The relevant question, therefore, is not a simple dichotomous “Which treatment is better?” approach but, given each patient’s unique and dynamic circumstance, “What is the best treatment option at that time and in that patient?” Important questions about the management of lumbar disc herniation such as the relative effectiveness of operative compared with nonoperative treatment for patients with an objective neurological deficit (such as numbness or foot drop), the effect of symptom duration and/or severity on treatment outcome, the relative risk of an adverse neurological outcome in nonoperatively treated patients, or the identification of prognostic factors predictive of treatment response, for example, could not be addressed in the SPORT, both because of the limitations in study design and the high crossover rate. In the end, the SPORT results offer no new data, insight, or conclusions into our understanding or potential management of lumbar disc herniation. Thus, it is difficult to comprehend Dr. Weinstein’s claim: “This study for the first time showed very significant improvement in the patients who ended up with non-operative treatment.” This decades old observation forms the very essence of contemporary management of lumbar disc herniation.

“Bias Toward the Null” SPORT Study Design

A major concern regarding the SPORT was that it was strongly biased toward the null hypothesis. In essence, due to such bias, the study would fail to identify a true difference between operative and nonoperative treatment for lumbar disc herniation when, in fact, a true difference exists. Indeed, the SPORT investigators’ finding that “im-

provements were consistently in favor of surgery at all time periods but were small and not statistically significant."

23 is compelling evidence that a Type II error occurred (that is, the study was underpowered). The three main reasons for this concern include: nonrepresentative aspects of the SPORT population, a failure to account for the anticipated high noncompliance in the study design, and a failure to stratify a priori treatment groups by well-established prognostic factors. All three factors are likely to have spuriously reduced the apparent treatment effect differences between operative and nonoperative treatment. One only has to compare the results obtained in the SPORT observational cohort and the as-treated SPORT RCT with the results obtained in the intent-to-treat RCT analysis to appreciate just how substantial this bias was. Although the first factor was probably not preventable, the latter two factors were inexplicably not controlled for or even considered in the study design of the SPORT, omissions that raise troubling questions about the integrity of this study. These three biased-toward-the-null concerns are discussed in detail.

Nonrepresentativeness of the SPORT RCT Study Population. There are two main concerns regarding the generalizability of the SPORT RCT. First, in contemporary clinical practice, operative and nonoperative management are not interchangeable treatments. Most patients presenting with symptomatic lumbar disc herniation initially undergo nonoperative care. Individuals with mild and/or improving symptoms continue with nonoperative management whereas those with continued severe, persistent, and/or worsening symptoms may choose a surgical option. In the SPORT RCT, however, many of the patients randomized to undergo surgery had relatively mild and/or improving symptoms. Indeed, nearly 20% of patients who accepted randomization into the SPORT RCT reported that their symptoms were “getting better” on baseline self-assessment. Many, if not most, of these patients would not be offered or would not choose surgery outside of the study. Further, compelling Class II evidence and accumulated empirical experience indicate that surgery in patients with mild symptoms tends not to have a significant treatment advantage compared with nonoperative care.1 Irrespective of whether these patients had surgery or crossed over to nonoperative treatment, their outcome would be credited to surgery. The true benefit of surgery would therefore be diluted by the inclusion of these patients in the SPORT RCT compared with those in contemporary clinical practice. The opposite is true for the patients with severe symptoms who were randomly allocated to receive nonoperative treatment. In contemporary clinical practice, many, if not most, of these patients would be treated operatively because they have been shown to derive the greatest differential benefit from surgery.3 These patients did preferentially cross over to surgery in the SPORT RCT, but the surgical benefit accrued to the nonoperative treatment arm because of the intent-to-treat analysis. In both circumstances, the net effect was to underestimate the true benefit of surgery and overestimate the benefit of nonoperative management within the SPORT RCT compared with contemporary clinical practice.

Secondly, if, as just noted, patients with severe symptoms tend to derive the greatest benefit from surgery and if patients with severe symptoms were less likely to enroll in the SPORT RCT, this would have the effect of diluting the true benefit of surgery within the SPORT compared with contemporary clinical practice. A review of the observational SPORT population confirms this suspicion.24 Indeed, the 719 patients in the SPORT observational cohort consisted of two distinct populations. Nearly 75% of these patients had severe symptoms—substantially worse than the average symptom severity in the SPORT RCT—and chose surgery, whereas approximately 25% of patients had relatively mild symptoms—substantially less than the average symptom severity in the SPORT RCT—and initially elected nonoperative treatment, although 22% ultimately crossed over and underwent surgery. Thus, severely affected patients were much more likely, by a margin of three to one, to refuse participation in the RCT. For the SPORT investigators to simply combine these two distinct populations from the observational cohort and demonstrate similarity between their averaged scores with the RCT population scores (see Table 1 of the study) is a specious comparison. It is likely that a similar ratio existed for the 771 patients who refused both randomization and participation in the SPORT observational cohort. The net effect of this uneven participation in the SPORT RCT would be to dilute further the true benefit of surgery within the SPORT RCT compared with that seen in contemporary clinical practice.

Failure to Account for Anticipated High Crossover Rate in the Study Design. The SPORT investigators’ admission that the SPORT RCT “did not account for any specific levels of nonadherence” is extremely troubling and likely doomed this study to fail in validly assessing its only intended hypothesis. The SPORT investigators estimated an 85% probability (that is, power) “to detect a 10-point difference in the SF-36 Bodily Pain and Physical Function scales or a similar effect size in the Oswestry Disability Index changes (i.e. effect size)” with their chosen sample size. There was no doubt, however, that a substantial crossover rate would occur in the SPORT RCT as it consistently has in previous studies of lumbar disc herniation and spinal stenosis.1,3,21 The net effect of such noncompliance is to reduce the magnitude of the difference in outcomes between the two groups. Patients with severe symptoms who were randomized to receive nonoperative treatment were more likely to cross over to surgery whereas patients with milder symptom severity who were randomized to undergo surgery were most likely to cross over to nonoperative treatment. Such crossover would clearly not only reduce the ability of the study to detect a true difference between the two treatments (that is, the study’s power) when, in fact, a true difference existed, but it would also virtually eliminate any possibility that surgery could ever show a benefit that exceeded a minimally clinically important difference threshold in the SPORT study design. Although increasing the number of study patients in the SPORT RCT may have allowed the identification of a statistically significant difference between treatment outcomes, no plausible level of increase in study subjects would have allowed the finding of a minimally clinically important difference between nonoperative and operative groups because of the high rate of crossover. By failing to incorporate any crossover rate into their study design or power calculation, the effective power of the SPORT RCT, based on the greater than 40% actual crossover rate, was only 7%, not the stated 85% (Fig. 1). Even had the SPORT investigators assumed the mini-
Results of the SPORT for lumbar disc herniation

The extraordinarily high crossover rate in the SPORT RCT precludes any valid comparison between actual operative and actual nonoperative treatment for painful lumbar disc herniation, the only a priori hypothesis of the SPORT. It also completely undermines any comparison between as-treated groups and prohibits any attempt to generalize the findings of the nonoperative group to all patients with a herniated lumbar disc. The reasons are clear. Random patient allocation ensures comparability only between the initial treatment assignment groups. Both initially randomized groups, therefore, were also equally representative of the overall population of patients with lumbar disc herniation from which the SPORT patients was sampled, although, as previously noted, patients with more severe symptoms were more likely to not participate in the SPORT RCT. It is important to realize, however, that the randomly allocated treatment assignment groups at the beginning of the RCT were quite different from the actual treatment (that is, as-treated) groups at the conclusion of the study. With each patient crossover, for example, the initial nonoperative treatment assignment group and the ultimate as-treated

failure to stratify prognostic factors a priori. As previously noted, lumbar disc herniation is a heterogeneous condition with respect to treatment response. Indeed, in the Maine Lumbar Spine Study, pretreatment symptom severity strongly correlated with treatment outcome. Patients with severe sciatica experienced clinically important and statistically significant superior outcomes after undergoing surgery compared with nonoperative management that persisted throughout the study period. In patients with mild symptoms, however, there was no benefit to operative compared with nonoperative management. That symptom severity, among other possible variables, is a prognostic factor was well known to the SPORT investigators. Non-stratified random allocation of known prognostic variables does not clarify individual treatment considerations, as it does with confounding and bias but, quite the contrary, it obscures them because the beneficial effect of surgery in one distinct subgroup of patients (such as those with severe symptoms) will be effectively diluted or nullified by the lack of benefit in another subgroup (such as those with mild symptoms). Unless these subgroups are identified a priori, little credence will be given to their post hoc identification and analysis.

Data Analysis and Presentation

It is clear that the SPORT results have been misrepresented and misinterpreted by many in the lay press and medical community. The misinterpretation by the lay press may be related, in part, to the methods of data analysis and the manner of data presentation. For example, as previously noted, the investigators provided virtually no analysis that took into account the high treatment failure rate in both the observational cohort (22%) and RCT (45%) nonoperative groups. Furthermore, despite the fact that the investigators clearly stated that direct treatment comparisons were not valid based on the intent-to-treat analysis, such direct (intent-to-treat) comparisons appeared extensively in textual, tabular, and graphic forms throughout the JAMA publication whereas alternative data analysis (for example, as-treated crossover censure analyses) was either extremely limited or completely absent.

The meaning of the ambiguous term “treatment group” throughout the SPORT RCT (“Patients in both the surgery and the nonoperative treatment groups improved substantially over a 2-year period” and “As shown in FIGURE 2, both treatment groups showed strong improvement . . .”) was obviously misinterpreted by lay press and the medical community to be synonymous with an actual treatment group. A more accurate categorical term such as “initial nonoperative (or operative) treatment assignment group” would have been preferable, particularly with knowledge of the high crossover rate. In all, one could infer that the manner of data analysis and presentation may have been, at best, imbalanced and confusing and, at worst, misleading. It is not clear whether the methods of data analysis and presentation were the choice of the SPORT investigators or dictated by the JAMA editorial board.

The SPORT Conclusions, Interpretations, and Extrapolations

The extraordinarily high crossover rate in the SPORT RCT precludes any valid comparison between actual operative and actual nonoperative treatment for painful lumbar disc herniation, the only a priori hypothesis of the SPORT. It also completely undermines any comparison between as-treated groups and prohibits any attempt to generalize the findings of the nonoperative group to all patients with a herniated lumbar disc. The reasons are clear. Random patient allocation ensures comparability only between the initial treatment assignment groups. Both initially randomized groups, therefore, were also equally representative of the overall population of patients with lumbar disc herniation from which the SPORT patients was sampled, although, as previously noted, patients with more severe symptoms were more likely to not participate in the SPORT RCT. It is important to realize, however, that the randomly allocated treatment assignment groups at the beginning of the RCT were quite different from the actual treatment (that is, as-treated) groups at the conclusion of the study. With each patient crossover, for example, the initial nonoperative treatment assignment group and the ultimate as-treated
nonoperative treatment group became less comparable with each other and, therefore, the as-treated nonoperative treatment group became less representative of the overall population of patients with lumbar disc herniation. Therefore, it is not possible to extrapolate any findings from the self-selected as-treated nonoperative treatment group to all patients with the diagnosis of lumbar disc herniation. To infer comparability or inevitability of outcome, irrespective of treatment received, simply from extrapolated data from the self-selected as-treated nonoperative treatment group is invalid. Similarly, any suggestion that the actual as-treated surgery patients would have done as well without surgery if they had just “waited a while longer” is also unjustified. Thus, the definitive statement by Dr. John Loesser that “All an operation does is shorten the duration of the pain. If you can do it with pain meds, the patient will be better within two years without surgical risks,” is simply not a valid conclusion based on the SPORT results.

In addition, the observation that nonoperative treatment was associated with “substantial improvement” in both the SPORT observational cohort and RCT is somewhat specious and reflects a fundamental flaw of both SPORT studies’ formulations. It is virtually impossible in contemporary clinical practice or in either arm of the SPORT study, for example, to have any outcome other than substantial improvement for the cases of nonoperative treatment of lumbar disc herniation. In contemporary clinical practice, for example, nearly all patients presenting with painful lumbar disc herniation will elect to have surgery. Thus, an analysis of data obtained in patients who only undergo nonoperative treatment will include predominantly those self-selected patients who were satisfied with their care and outcome, because in most cases of nonoperative treatment failures patients eventually undergo surgery. Conversely, patients who have undergone surgery, usually after nonoperative treatment fails to improve symptoms, can never cross over to another treatment, no matter their outcome or satisfaction. In effect, most patients who undergo surgery are those in whom nonoperative treatment has failed. The implication is that patients with poor treatment prognoses (for example, chronic pain, certain medical and psychosocial comorbidities, irreparable nerve injury, and diagnostic uncertainty) will be disproportionately represented in any as-treated surgical group. In the end, cases of surgical treatment failure persist whereas patients in cases of failed nonoperative treatment simply cross over.

In the SPORT observational cohort, only the patients who stayed in the nonoperative group throughout the 2-year study period were included in the nonoperative treatment group analysis. The 22% of patients in whom nonoperative treatment failed and who crossed over to the surgery arm were analyzed in the surgical group. Despite the failure of the SPORT investigators to account for this 22% rate of nonoperative failure in their analysis, the foreordained improvement in the nonoperative group was significantly inferior to outcomes in the surgically treated group at every follow-up interval throughout the study. This treatment effect difference remained substantial, even taking into account the SPORT investigators’ decision to adjust for (that is, to average) the significant differences in preoperative pain and disability scores between the two treatment groups. The net effect of this adjustment was to lower the absolute magnitude of clinical improvement with surgery and to elevate the absolute magnitude of the nonoperative treatment effect.

The SPORT RCT was even more problematic with respect to nonoperative treatment failures. Not only was the nonoperative treatment failure rate substantially higher (45%) in the RCT, but the investigators did not analyze these treatment failures either through their categorical identification or censure. Worse yet, the patients in whom nonoperative treatment failed who crossed over and, presumably, experienced a successful outcome with surgery actually contributed more to the magnitude of benefit credited to nonoperative treatment than cases in which the patients who actually stayed in the nonoperative treatment group throughout the study period, according to the intent-to-treat analysis.

There are additional implications. For example, the fact that there were no instances of severe neurological deterioration from a cauda equina syndrome among the self-selected as-treated nonoperative treatment group cannot be generalized to all patients with symptomatic lumbar disc herniation, particularly because the more severely symptomatic surgically treated patients would intuitively be more likely at risk for adverse neurological outcomes. How many of the patients in the SPORT RCT with severe symptoms who chose or were assigned to surgery avoided or resolved an adverse neurological outcome? How many of the 45% of patients who were initially assigned to nonoperative treatment in the RCT but crossed over to surgery avoided or experienced resolution of an adverse neurological outcome? The same question applies to the SPORT observational cohort, both for the surgically treated patients and the 22% of initially nonoperatively treated patients who crossed over to surgery. These questions are also relevant to the 771 patients who refused to be in either the SPORT observational cohort or RCT.

Furthermore, because of patient noncompliance and limitations in SPORT study design, we cannot determine from either SPORT study if there was a difference in the efficacy of early compared with late surgery or surgery compared with no surgery on either symptom relief, such as chronic radicular pain, or recovery from an established neurological deficit, such as asymmetrical weakness (for example, foot drop) or numbness, a much more common occurrence than a cauda equina syndrome. In reality, the isolated observation that no patient in the self-selected as-treated nonoperative group developed a cauda equina syndrome provides little insight into the potential deleterious effects of time for two reasons. First, a delayed-onset cauda equina syndrome is extraordinarily rare, particularly in the patients with relatively less severe and/or improving symptoms who ultimately comprised the as-treated nonoperative group. Second, far more common adverse outcomes were not even measured or analyzed postoperatively (for example, foot drop, radicular weakness, or numbness) in either SPORT study. Thus, Dr. Weinstein’s conclusion that “… nobody got worse. We never knew that until we did the study” is neither broadly applicable nor completely accurate. Dr. Carragee’s statement in the accompanying JAMA editorial that “the SPORT data clearly show that the risk of serious prob-
Results of the SPORT for lumbar disc herniation

...lems (neurolologic deterioration, cauda equina syndrome, or progression of spinal instability) when receiving nonoperative care is extremely small. The fear of many patients and surgeons that not removing a large disk herniation will likely have catastrophic neurolologic consequences is simply not borne out" is only applicable to the highly self-selected as-treated nonoperatively managed cases.

A long-standing neurolologic deficit may not resolve as readily or completely as a subacute condition. Although not as devastating as a cauda equina syndrome, permanent leg weakness and/or numbness can be a substantial lifelong disability. Time, therefore, may not always be an ally but can represent a double-edged sword. Thus, a broadly generalized no-harm-in-waiting conclusion, derived only from the outcome of the self-selected nonrepresentative as-treated nonoperative treatment group, is simply not valid.

Conclusions

The SPORT observational cohort study reflects just how efficient and rational patient choice is with respect to the contemporary management of painful lumbar disc herniation. Patients with severe, unrelenting, or worsening pain, and/or neurological deficit, typically chose surgery and had excellent outcomes whereas most patients with less pain and disability usually chose nonoperative treatment. Although most of these nonoperatively treated patients also had their symptoms improve, 22% with persistently more severe symptoms ultimately crossed over to undergo surgery. Analysis of the SPORT RCT data is equally informative. Patients who underwent surgery, whether in assigned or crossover cases in the RCT, shared remarkably similar preoperative pain and disability scores. The same was true for the patients who, by assignment or crossover, received nonoperative treatment. In both the SPORT observational cohort and the RCT, therefore, baseline values were substantially worse in virtually every category including pain, physical function, disability, neurological function, and perception of worsening in the patients who actually underwent surgery than those who ultimately underwent nonoperative therapy. Thus, although treatment allocation in the SPORT RCT was random, the actual treatment received was essentially stratified according to disease severity by patient choice. In the end, rational patient choice effectively nullified the SPORT randomization process. To even perform an intent-to-treat analysis under circumstances in which the SPORT RCT so closely approximated the observational cohort, with patients feeling free to make their own treatment choices, makes little sense.

For non–life-threatening, heterogeneous conditions in which patient preferences are appropriately a major determinant of treatment, randomized studies are extraordinarily difficult to conduct, as the experience of the SPORT investigators demonstrates. These studies did, however, reaffirm our confidence that informed patients are able to determine the best treatment for them. Although prospective RCTs in which investigators use an intent-to-treat analysis may be methodologically valid, they guarantee neither truth nor clarity, as is vividly demonstrated in the SPORT RCT for lumbar disc herniation. Quite the contrary, the SPORT study and its interpretation contributed to greater confusion regarding the evaluation and management of patients with lumbar disc herniation, while adding no new data, insight, or perspective. I only hope the investigators, the medical community (including peer-reviewed journals), and the lay press will be more responsible in their analysis, interpretation, and reporting of the upcoming SPORT results on spinal stenosis and degenerative spondylolisthesis.

Acknowledgments


References


Samson K: Major trial finds surgery helps disabling sciatica, but non-surgical patients also improve over time. Neurology Today 6:16–17, 2006

Springen K: To cut or not to cut, in Newsweek: November 21, 2006 (http://www.msnbc.msn.com/id/15840801/site/newsweek/) [Accessed 22 March 2007]


Accepted February 23, 2007.

Address reprint requests to: Paul C. McCormick, M.D., M.P.H., Department of Neurosurgery, Columbia University College of Physicians and Surgeons, 710 West 168th Street, 5th Floor, New York, New York 10032. email: pcm5@columbia.edu.