

Design of the Spine Patient Outcomes Research Trial (SPORT)

Nancy J. O. Birkmeyer, PhD,* James N. Weinstein, DO, MS,*† §
Anna N. A. Tosteson, ScD,†‡ Tor D. Tosteson, ScD,‡ Jonathan S. Skinner, PhD,||
Jon D. Lurie, MD, MS,† Richard Deyo, MD, MPH,¶ and John E. Wennberg, MD, MPH‡

Summary of Background Data. The Spine Patient Outcomes Research Trial (SPORT) was designed to assess the relative efficacy and cost-effectiveness of surgical and nonsurgical approaches to the treatment of common conditions associated with low back and leg pain.

Objectives. To describe the rationale and design of the SPORT project and to discuss its strengths and limitations.

Study Design. Descriptive.

Methods. First, the authors explain the rationale for embarking on SPORT, *i.e.*, deficiencies in the existing scientific knowledge base for treatment of these conditions. Second, the authors describe the design of SPORT, including topics such as specific aims, participating sites, study population, recruitment and enrollment, study interventions, follow-up, outcomes, statistical analysis, and study governance and organization. Finally, issues that complicate the performance of randomized trials in surgery as they relate to the design and conduct of SPORT are discussed.

Results. The SPORT project is being conducted at 11 clinical centers around the United States. It involves the simultaneous conduct of three multicenter, randomized, controlled clinical trials. The study includes patients with the three most common diagnoses for which spine surgery is performed: intervertebral disc herniation, spinal stenosis, and degenerative spondylolisthesis, and it compares the most commonly used standard surgical and nonsurgical treatments for patients with these diagnoses. By the end of enrollment the authors anticipate a total of 500 patients with intervertebral disc herniation, 370 patients with spinal stenosis, and 300 patients with degenerative spondylolisthesis in the randomized trials. Patients who meet the eligibility criteria but decline to be randomized are invited to participate in an observational cohort study. Patients are being followed for a minimum of 24 months with visits scheduled at 6 weeks and at 3, 6, 12, and 24 months.

Conclusions. The results of this study will provide high-quality scientific evidence to aid clinical decision-making and improve treatment outcomes for these common, costly, and, in some instances, debilitating conditions. [Key words: randomized controlled trial, observational cohort, intervertebral herniated disc, spinal stenosis, low back pain, sciatica, degenerative spondylolisthesis, outcomes, cost-effectiveness] **Spine 2002; 27:1361–1372**

Few areas of clinical medicine are as controversial as the management of conditions associated with low back pain, as evidenced by wide variability in the use of spine surgery both nationally and internationally.^{7,52} Overall rates of spine surgery in the United States have increased dramatically in recent years.⁶ Between 1988 and 1997, spine surgery rates among U.S. Medicare enrollees grew by 57%, from 2.1 to 3.4 per 1000 population.⁶ Nearly half a million cases of spine surgery were performed in the United States in 1999.¹⁴ Although only about 10% of patients with low back pain have the conditions that are the focus of the Spine Patient Outcomes Research Trial (SPORT) (intervertebral disc herniation [IDH], spinal stenosis [SpS], and degenerative spondylolisthesis [DS]),¹⁰ these conditions are especially important to study because they account for a large proportion of the spine surgery that is performed in the United States each year.¹⁰

The lack of clinical consensus regarding the use of spine surgery can be traced to deficiencies in the scientific data that underlie indications for its use.^{11,12} There are a few cohort studies comparing surgical and nonsurgical treatment for patients with the conditions that are the subject of SPORT.^{1–4,22,25,38,42} However, scientific flaws inherent to cohort studies, especially large baseline differences in the treatment groups resulting from selection bias, preclude definitive conclusions regarding the efficacy of surgical and nonsurgical treatment from these data. There is also one randomized, controlled study of surgical and nonsurgical treatment for patients with herniated disc.⁴⁹ This study also had some important flaws. In particular, the primary measure of patient outcomes was the investigator's categorization of patients' verbal assessments of the results of treatment. The extent to which this measure correlates with patients' self-evaluations of clinical, functional, or satisfaction outcomes is unknown. In addition, both surgical and nonsurgical treatments have changed dramatically since the

From the Departments of *Surgery, †Medicine, ‡Community and Family Medicine, and the §Clinical Trials Center, Dartmouth Medical School, Lebanon, and the ||Department of Economics, Dartmouth College, Hanover, New Hampshire; and the ¶Departments of Medicine, and Health Services, and the Center for Cost and Outcomes Research, University of Washington, Seattle, Washington.

Supported by grants from the National Institutes of Health (National Institute of Arthritis and Musculoskeletal and Skin Diseases and Office of Research on Women's Health, the Centers for Disease Control and Prevention), National Institute of Occupational Safety and Health (U01-AR45444-01A1), and Agency For Healthcare Quality and Research (K02 HS11288).

Acknowledgment date: August 10, 2001.

Acceptance date: October 15, 2001.

The manuscript submitted does not contain information about medical device(s)/drug(s).

Federal funds were received to support this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

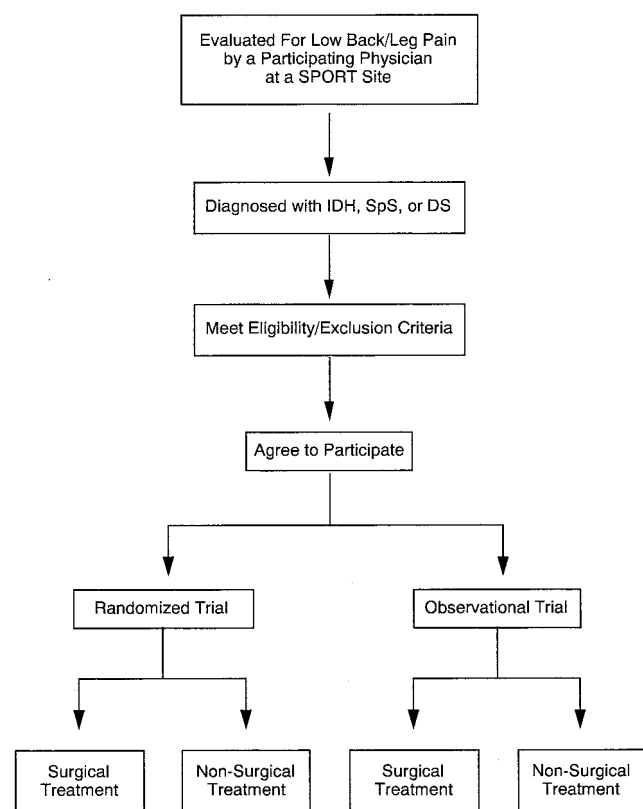


Figure 1. Overview of the SPORT study design.

time the study was conducted (patients enrolled from 1970 to 1971). The rest of the literature on the treatment of these conditions is based on uncontrolled case series. Although such studies may be valuable as an initial assessment of treatment safety, they are not adequate to demonstrate effectiveness. This is particularly true for conditions such as these in which spontaneous improvement may result in biased estimates from uncontrolled studies.

Research addressing the causes and treatment of spinal conditions has been given high priority because of its prevalence and associated morbidity, costs, and clinical controversy. The SPORT project application was fortunate to be successful in obtaining funding from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) of the National Institutes of Health (NIH). The purpose of this article is to describe the de-

sign of SPORT (Figure 1) and discuss some of its strengths and limitations.

■ Study Design

Study Aims.

SPORT has three specific aims:

1. To simultaneously conduct three multicenter randomized controlled trials comparing surgical and non-surgical treatment for patients with IDH, SpS, or DS that are considered eligible for surgery with repeated longitudinal measurement up to 24 months.

2. To characterize subjects who decline participation in the randomized controlled trial but agree to be followed as part of an observational cohort. For these subjects we record all treatments, outcomes, and costs.

3. To formally estimate the cost-effectiveness of surgical *versus* nonsurgical interventions for IDH, SpS, and DS through a synthesis of the results from the randomized controlled trial and the observational study cohorts.

Study Sites.

The 11 clinical centers that recruit patients into SPORT include the following: Dartmouth–Hitchcock Medical Center, Hanover, NH; Rothman Institute at Thomas Jefferson Hospital, Philadelphia, PA; Rush–Presbyterian–St. Luke’s Medical Center, Chicago, IL; Emory University Medical Center, Decatur, GA; University Hospitals of Cleveland/Case Western Reserve, Cleveland, OH; New York University/Hospital for Joint Diseases, New York, NY; William Beaumont Hospital, Royal Oak, MI; University of California Medical Center–San Francisco, San Francisco, CA; Washington University Medical Center, St. Louis, MO; Nebraska Foundation for Spinal Research, Omaha, NE; and the Hospital for Special Surgery, New York, NY. Each of these sites is a member of the National Spine Network. This nonprofit organization was formed in 1994 to develop and foster high-quality and cost-effective care for its member’s patients. To this end the organization developed an infrastructure for collecting outcomes data and a registry of patients with spinal conditions that provided the pilot data for the SPORT proposal.

Study Population.

SPORT inclusion (Table 1) and exclusion criteria were developed from evidence-based management algorithms

Table 1. Inclusion Criteria by Diagnosis

Inclusion Criteria	IDH	SpS	DS
Physical examination	Radicular pain and evidence of nerve root compression	Pseudoclaudication or radicular pain with associated neurologic deficit	Pseudoclaudication or radicular pain with associated neurologic deficit
MRI	Herniated disc at level/side of symptoms	Lumbar spinal stenosis at one or more levels	Spinal stenosis at the level of listhesis (L3–L4 or L4–L5) with or without adjacent level stenosis
Plain films	No radiographic inclusion criteria	No radiographic inclusion criteria	Degenerative forward slippage of the L3 relative to the L4 vertebral body or the L4 vertebral body relative to the L5 vertebral body in the sagittal plane on lateral view

produced by the Agency for Health Care Policy and Research⁸ and others¹⁹ as well as by consensus of the participating physicians.

Inclusion Criteria. For IDH, patients are eligible for SPORT if they have radicular pain and evidence of nerve root compression with a positive nerve root tension sign (positive straight leg raise test or femoral tension sign). Alternatively, they may have a reflex (asymmetric depressed reflex), sensory (asymmetric decreased sensation in a dermatomal distribution), or motor (asymmetric weakness in a myotomal distribution) deficit with associated radicular symptoms and positive nerve root tension signs. In addition, a confirmatory imaging study (MRI or CT) must indicate an IDH (a protrusion, extrusion, or sequestered fragment) at a location (level and side) corresponding to the patient's radicular signs or symptoms.¹⁹ Patients with only a bulging disc (circumferential symmetric extension beyond the interspace) are not eligible.¹⁹

Patients are eligible for the SpS trial in SPORT if they have pseudoclaudication (pain in the buttock, thigh, or leg on ambulation that improves with rest) or radicular pain with an associated neurologic deficit. The patient with suspected pseudoclaudication must have strong distal pulses or other testing to rule out vascular claudication. In addition, SpS patients must have a confirmatory imaging study (MRI or CT) showing lumbar SpS at one or more levels (L2 to sacrum) defined as narrowing of the central spinal canal, lateral recesses, or neural foramina due to encroachment on the neural structures by the surrounding bone and soft tissue. Patients are not eligible if they have evidence of instability on lateral flexion-extension radiographs, defined as a change of $>10^\circ$ of angulation of adjacent segments by Cobb measurement or a change of more than 4 mm of anteroposterior or posteroanterior translation.

To be eligible for the DS trial in SPORT, patients must have pseudoclaudication or radicular pain with an associated neurologic deficit and have a confirmatory imaging study (MRI or CT) indicating SpS at L3–L4 or L4–L5 and a degenerative forward slippage (of the L3 relative to the L4 vertebral body or the L4 vertebral body relative to the L5 vertebral body) in the sagittal plane on lateral standing radiograph. Displacement is usually $\leq 25\%$ of the vertebral body. Patients with adjacent levels of stenosis are eligible (additional levels of stenosis are acceptable at L2–L3 and/or L4–L5 for degenerative slips at L3–L4 and at L3–L4 and/or L5–S1 for degenerative slips at L4–L5).

Exclusion Criteria. Patients in all three diagnostic categories are not eligible for SPORT if they have any of the following exclusion criteria:

- Insufficient trial of nonsurgical treatment (6 weeks for IDH and 12 weeks for SpS or DS);
- Cauda equina syndrome or progressive neurologic deficit requiring urgent surgery;

- Overall health that makes spine surgery too life-threatening to be appropriate;
- Dramatic improvement with nonsurgical care;
- Possible pregnancy;
- Active malignancy (patients with a history of any invasive malignancy except for nonmelanoma skin cancer must have been treated with curative intent and have been free from clinical signs and symptoms of the malignancy for at least 5 years);
- A current fracture, infection, or significant deformity ($>15^\circ$ lumbar scoliosis) of the spine;
- <18 years of age;
- Prior lumbar spine surgery;
- Current enrollment in another experimental spine-related protocol; or
- Not available for follow-up or unable to complete questionnaires.

Recruitment and Enrollment

A research nurse and the participating physicians at each site identify potential subjects from among their clinic's patients. In addition, SPORT has been advertised on a national level *via* press releases issued by the NIH/NIAMS and through guest speaking appearances by the principal investigator at various specialty society meetings. The participating clinical sites have advertised the project on a local level *via* print, television, and radio media. SPORT is also publicized on the NIH/NIAMS web site and on its own web site at <http://sport.dartmouth.edu/nsn>.

Videotapes are used as part of the recruitment process to standardize and make available information that potential subjects need to make an informed decision about participation. Evidence-based programs produced by the Foundation for Informed Decision Making, "Treatment Choices for Low Back Pain: Herniated Disc" and "Treatment Choices for Low Back Pain: Spinal Stenosis," were edited to include information about SPORT. The programs provide information from the published literature regarding what currently is and is not known about the risks and benefits of the various treatment options for these conditions in lay language.

Informed consent is obtained from patients who are willing and eligible to participate. Data regarding demographic characteristics, medical history and comorbidities, symptoms (bothersomeness and frequency), and baseline measurements for all outcomes are obtained *via* a combination of patient interview, patient self-administered survey, and physician survey. All data are collected *via* portable computers that are equipped with software for touch-screen data entry or *via* a SPORT Internet site developed for this project.

Randomized Trials

Patients who agree to participate in a randomized arm of the trial receive their treatment assignment at the time of enrollment. To assure that the number of subjects is approximately the same in the surgical and nonsurgical groups for each site/disease group combination at the end of accrual, with approximately balanced assignment

totals at interim intervals as well, a variable blocked randomization scheme is used.²⁰ “Blocked” randomization means that treatment assignment is performed sequentially within blocks of a predetermined size so that when a blocked sequence is completed it will contain nearly equal numbers from each treatment group. “Variably blocked” means that the sequence and size of the blocks are varied randomly to further assure that the treatment assignments cannot be predicted or manipulated.

An automated randomization system was created by the Statistical Center for SPORT based on computer-generated, random, blocked treatment assignments for each disease group within each site. The treatment assignment tables are accessed through either a web site maintained by the Statistical Center or through portable computers provided for each site with special software for administering questionnaires. When eligibility criteria have been verified, the program extracts the next randomization code from the appropriate disease sequence along with the sequence number and a confirmation code. These codes are permanently embedded in the patient record along with the exact time of randomization. Automated reports for enrollments are updated on demand through the secured Statistical Center web sites and allow coordinators and investigators to track enrollment performance on a daily basis.

Observational Cohorts

Patients may choose not to be randomized but still be part of the study by participating in the observational cohort study. The inclusion and exclusion criteria, study intervention protocols, follow-up schedule, and study endpoints are identical to patients in the randomized study, but treatment is not randomly assigned.

Study Interventions

Surgical

The participating physicians have agreed to use the following standardized surgical approaches by diagnosis:

- IDH—Decompression of the involved nerve root by a standard open or microdiscectomy;
- SpS—Standard posterior decompressive laminectomy;
- DS—Standard posterior decompressive laminectomy with or without an additional bilateral single-level fusion (autogenous iliac crest bone grafting at the level of the listhesis with or without posterior instrumentation using pedicle screws).

In addition, participating physicians agreed to forego the use of any experimental devices or biologics as part of these procedures. However, if the physician decides during the surgery that the patient requires a procedure that differs from these protocols, he or she is instructed to perform that preferred procedure and record the details.

Nonsurgical

At a minimum, patients in all three diagnostic groups receive active physical therapy, education/counseling

Table 2. Nonsurgical Treatment Options

Patient education/counseling
Back school
Home exercise instruction
Coping skills training
Vocational counseling
Psychological assessment and counseling
Passive physical therapy
Ultrasound/phonophoresis
Iontophoresis
Heat/ice
Massage
Mechanical traction
Active physical therapy
Flexibility
Strengthening
Conditioning
Stabilization
McKenzie
Maitland
Cyriax
Manipulation
Chiropractic
Osteopathic
Activity restriction
Bed rest (1–3 days)
Lifting restriction
Sitting restriction
Driving restriction
Work restriction
Devices
Braces
Shoe inserts
Corset
Cane
Walker
Injections
Epidural
Facet joint
Trigger point
Functional restoration program
Bundled program of physical therapy, occupational therapy, counseling, etc.
Alternative/nontraditional treatments
Acupuncture
Therapeutic touch
Interferential
Electrical stimulation techniques
Feldenkrais
Muscle energy
Reiki therapy
Energy work
Magnets

with home exercise instruction, and a nonsteroidal anti-inflammatory drug if tolerated. Patients may receive any additional nonsurgical therapies deemed appropriate by their physician with all prescribed therapies recorded. Participating sites are encouraged to aggressively use all appropriate nonsurgical therapies in those nonsurgical patients who do not respond to the minimum intervention.

A complete list of nonsurgical therapies and medications being tracked in SPORT are shown in Tables 2 and 3. These lists are meant to include both typical treatments that would be prescribed by the participating physician as well as treatments patients might seek on their own. This approach, as opposed to a more limited, standardized nonsurgical intervention in all patients, is more

Table 3. Medication Options

Nonprescription pain
Acetaminophen
Aspirin
Nonsteroidal anti-inflammatory drugs
Ibuprofen
Naproxen
Diflunisal
Salsalate
Diclofenac
Etodolac
Flurbiprofen
Indomethacin
Ketoprofen
Ketorolac
Nabumetone
Oxaprozin
Piroxicam
Sulindac
Tolmetin
Cox-2 inhibitors
Celecoxib
Rofecoxib
Oral steroids
Dexamethasone
Methylprednisolone
Prednisone
Narcotics
Codeine
Oxycodone
Hydrocodone
Hydromorphone
Morphine
Meperidine
Metadone
Muscle relaxants
Cyclobenzaprine
Carisoprodol
Diazepam
Methocarbamol
Orphenadrine
Tricyclic antidepressants
Amitriptyline
Desipramine
Doxepin
Imipramine
Nortriptyline
Other antidepressants
Fluoxetine
Paroxetine
Sertraline
Venlafaxine
Other medications
Carbamazepine
Gabapentin
Tramadol
Alternative/nontraditional medications
Herbal remedies
Homeopathic remedies
Glucosamine
Chondroitin sulfate
Vitamins
Nutritional supplements

generalizable and will allow for a better understanding of the broad range of interventions and their associated costs, being used in the care of patients with these diagnoses.

Follow-Up

Follow-up data are gathered at two short-term follow-up intervals and five long-term follow-up intervals. The

short-term follow-ups are at 6 weeks and 3 months from the time of treatment onset (surgery for the surgical group and the start of nonsurgical therapy, generally at the time of enrollment, for the nonsurgical group). The difference in timing the short-term follow-up visits for the surgical and nonsurgical treatment groups was necessitated by wide variability in the amount of time between randomization and surgical treatment. This occurs for a variety of reasons, including surgeons' schedules and case review for insurance and Workers' Compensation. To address the potential biases that these follow-up timing differences might create, we developed a complex set of follow-up timing rules that require additional follow-up for some surgical patients.

Long-term follow-up occurs at 6, 12, and 24 months, and, if time permits, at 36 and 48 months from the time of enrollment. One-hour clinic visits with the nurse coordinator are scheduled for all follow-ups. When study visits correspond to routine clinical care visits, the meeting with the nurse coordinator is in addition to the usual time scheduled with their physicians. When study visits occur outside the usual care, if possible the patient is also scheduled to meet with their physician at the beginning of this appointment for a brief visit.

Outcomes

The primary outcome measure for the SPORT project is health-related quality of life as measured using both generic and disease-specific instruments. Secondary endpoints, which are included to address the cost-effectiveness of surgical *versus* nonsurgical treatment, include preference-based measures of current health, resource utilization, and work status.

Primary Outcomes

SF-36 Health Status Questionnaire. We chose as our generic health status measure the SF-36 Health Status Questionnaire.⁴⁸ This instrument consists of 36 questions that can be aggregated to form eight subscales (physical function, mental health, general health perceptions, pain, role limitations—physical, role limitations—emotional, social functioning, and vitality) and two summary scales (physical and mental component scales). On each scale higher scores indicate better outcomes. Scores can be compared with published age- and sex-matched general population or disease-specific norms.

Oswestry Disability Index. For our disease-specific instrument we chose the Oswestry Disability Index.¹⁸ We are using a version of the questionnaire that consists of nine questions relating to limitations in performing the following activities during the past week: getting dressed, lifting, walking and running, sitting, standing, sleeping, social and recreational activities, traveling, and sexual activity. Each question has six graded responses that range from unlimited, pain-free activity to total incapacity resulting from pain. The sum of the nine responses is expressed as a percentage of the maximum score. In the

original scoring of the instrument higher scores represented greater disability.¹⁸ We reverse the coding of the responses so that higher scores indicate better outcomes for consistency with the interpretation of the SF-36 measures.

Secondary Outcomes

Preference-Based Measures of Current Health. To assess the economic value of low back pain treatment we use quality-adjusted life-years as our measure of effectiveness. To estimate quality-adjusted life-years, preference-based measures (*i.e.*, utilities) for current health are required. Utility values,^{40,44} which range from 0 (worst imaginable health) to 1 (best imaginable health) reflect the desirability of health outcomes. Utilities are important to consider because patients with the same degree of functional limitation or symptoms may feel very differently about life under those circumstances.³⁶

In SPORT, utility for current health is estimated using two preference classification systems: the 5-item Euro-QoL EQ-5D¹⁶ and the 15-item Health Utilities Index.⁴⁷ Both provide societal utility values that are appropriate for estimating quality-adjusted life-years. In addition, we obtain individual patient values for current health using a visual analogue scale.¹⁶

Resource Utilization and Cost. In our analysis we distinguish among direct inpatient costs, direct outpatient costs, and indirect costs. To estimate costs associated with surgical and nonsurgical care, we track resource utilization and work status at each follow-up period. The patient is our primary source for resource utilization and work history data. To enhance the quality of the recalled cost data, two approaches are used. First, we provide all patients with detailed diaries, which serve as memory aids, for tracking encounters with health care providers. Second, we minimize the period of recall. To capture costs for the first 3 months of study enrollment, cost data are obtained at 6 weeks and 3 months using a 6-week recall window. At the 6-, 12-, 24-, 36-, and 48-month visits, we record most costs using a 1-month recall period and use it to estimate average costs for the months between visits. However, certain costs, including hospitalizations and medical devices, are captured for the entire period between study visits.

Direct Inpatient Costs. To identify resource use associated with hospitalization, patients are questioned at follow-up visits about hospitalizations since their last study visit. Inpatient services include both physician and hospital service components of costs for each of the hospital stays experienced during the period of follow-up. Additional costs incurred during hospitalization include physician and diagnostic service expenses. We estimate costs for these using the Resource-Based Relative Value Scale,²³ which is used by the Centers for Medicare and Medicaid Services.⁹ We favor this approach over using institution and cost-enter-specific cost-to-charge ratios

because it reflects national fee schedules and can therefore be generalized to the U.S. population.

Direct Outpatient Costs. To identify care other than hospitalization, patients are questioned at each follow-up visit about their use of outpatient services, medications, and devices. We collect the relevant physician costs for outpatient services using the same resource-based relative value scale weights as for the physician inpatient services. To estimate costs of medications and other health care services not covered by Medicare, we are surveying providers on charges for services at diverse geographic locations involved in this study. Cost estimates of physical therapy at each site are derived based on condition-specific profiles, which are being developed through surveys of each site's physical therapy practices.

Indirect Costs. An important component of the overall economic evaluation of surgical and nonsurgical treatment for these conditions is the measurement of indirect costs of employees, employers, and the government resulting from the spine-related condition. To this end our survey asks about occupation, difficulty performing work (either in a job or as a homemaker), job changes, and financial distress stemming from the spine condition. To estimate indirect costs associated with time lost from work, we collect earning and wage information at baseline to allow for assigning costs to lost work hours. Average weekly hours are collected at each study visit, whereas wage, insurance coverage, disability payments, and measures of difficulty working are collected at 6 months, 12 months, and yearly thereafter. Finally, we collect information about factors such as insurance status, disability payments, and income of other family members to determine what factors affect labor force participation and treatment options. We use the pretreatment wage rate in these calculations but also include a question at the 12-month follow-up regarding the post-treatment wage rate.

Monitored Events

A monitored event form is completed whenever it is learned that a study patient has been lost to follow-up, has crossed over from their randomly assigned or self-designated treatment group, or has withdrawn from the study. Deaths and hospitalizations are immediately reported to the study Principal Investigator, Data and Safety Monitoring Board, the NIAMS Project Officer, and all participating institutional review boards (IRBs). The Data and Safety Monitoring Board uses narrative data provided by the site regarding the circumstances of the event as well as any documentation in the medical record, discharge or operative notes, or death certificates to determine whether the event was related to the patient's spine condition or its treatment. All other medical complications are monitored throughout the study at all routine follow-ups and are reported to the Data Safety Monitoring Board and the NIAMS Project Officer at regular intervals. IRB notification

guidelines are followed according to each site's specifications as well as complying with national regulations (45CFR46, see <http://ohrp.osophs.dhhs.gov/humansubjects/guidance/45cfr46.htm>).

Protocol Violations

Any deviation from the study protocol is considered a protocol violation. Major protocol violations include the following: randomization of an ineligible patient, enrollment of a SPORT participant in any other spine-related study, study participant receiving the wrong treatment, randomization before completion of necessary paperwork or diagnostic studies, loss of radiology or operative report, and informed consent violations. Minor protocol violations include failure to report a monitored event within 24 hours of learning of it and failure to obtain follow-up data within the specified time intervals. Violations are reviewed weekly and serve as data for reports to the Data Safety Monitoring Board, NIAMS, and IRBs.

Statistical Analysis

Our primary analyses will conform to the "intent-to-treat" or "as randomized" principle. Thus, patients who cross over (elect surgery after having been randomized to the nonsurgical arm or decline to undergo surgery after having been randomized to that treatment) will be included in the treatment group to which they were randomized for the purposes of assessing the intervention. The intent-to-treat analysis is an estimate of the intervention effect attributable to the randomization event, *i.e.*, an estimate of the benefit of offering an intervention without regard to compliance. In this sense it is a more realistic assessment of the "effectiveness" of an intervention in a clinical setting and is statistically unbiased. Secondary analyses will examine the potential effects of treatment in the absence of crossover by tracking patient outcomes and actual treatment status.

All the data analyses required to fulfill the specific aims of the study will be accomplished using statistical methods for repeated measures and longitudinal data.²⁸ The primary study endpoints will be summarized in terms of changes from preintervention baseline quality of life measurements. The mean changes observed at specified follow-up times will be compared between the treatment arms with appropriate adjustment for intrapatient correlations. In the event of unexpected baseline imbalances between the treatment arms, adjusted analyses will include important predictors of quality of life scores as covariates.

Sample size calculations were based on our primary outcome measures for health-related quality of life, the SF-36 and Oswestry instruments. For the SF-36 calculations the physical function and bodily pain subscales were used. The required sample size for each trial was computed so that a *t* test based on the time-specific treatment effect with a two-sided significance level of 0.05 would have a power of 0.85 to detect mean differences from baseline to follow-up of at least 10 points, assuming a 20% loss to follow-up. Appropriate variances for

the quality of life scores within each disease group were derived from data provided by the National Spine Network.

Study Governance and Organization

The overall structure of SPORT (Figure 2) is similar to many large NIH-sponsored collaborative clinical trials. An Executive Committee, the membership of which includes the study's Principal Investigator, representatives of NIAMS, and a physician representative from the participating sites, oversees the study. The study is operationally directed by the Scientific Work Group, comprised of the Principal and Dartmouth Co-Principal Investigators as well as the study coordinator and a representative of the Site Principal Investigators. The Scientific Work Group meets regularly and as needed to attend to issues that arise throughout the various phases of the project, including protocol development and/or modification, enrollment, follow-up, data analysis, and reporting.

Data and Safety Monitoring Board

The Data and Safety Monitoring Board is composed of five voting members from outside the study investigative group with experience in statistical and scientific issues relevant to the conduct of large, collaborative clinical trials. This committee reviews summaries of safety, the accrual and progress of the trial, the quality of the data, and blinded interim efficacy/effectiveness analyses, and reports on its findings to the NIAMS Project Officer. It is the responsibility of the Data and Safety Monitoring Board to interpret data on adverse effects. This group reviews the data every 6 months and makes recommendations to NIAMS regarding actions to ensure that subjects are not exposed to undue risks. The mandate of the committee complies with the July 1, 1999 release of the NIH Policy for Data and Safety Monitoring, with its primary function being to "ensure the safety of participants and the validity and integrity of the data."

Study Coordinating Center

The Study Coordinating Center for SPORT (Figure 2) is located at Dartmouth Medical School/Dartmouth-Hitchcock Medical Center in Lebanon, NH. Operationally, there are four groups within the Study Coordinating Center: Statistics, Data Management, Study Administration, and Cost-Effectiveness.

Staff from the Data Management group preinstalled all software used in this study, taught participating center staff how to use it, and maintain ongoing communication with the sites to troubleshoot software and hardware problems.

Baseline, enrollment, treatment, and follow-up data are collected at the participating centers by a combination of patient self-report, physician report, or nurse coordinator interview (personal and phone). These data are entered electronically using touch-screen computers and secure web sites by patients and study staff at the participating centers. The data collection system has numerous

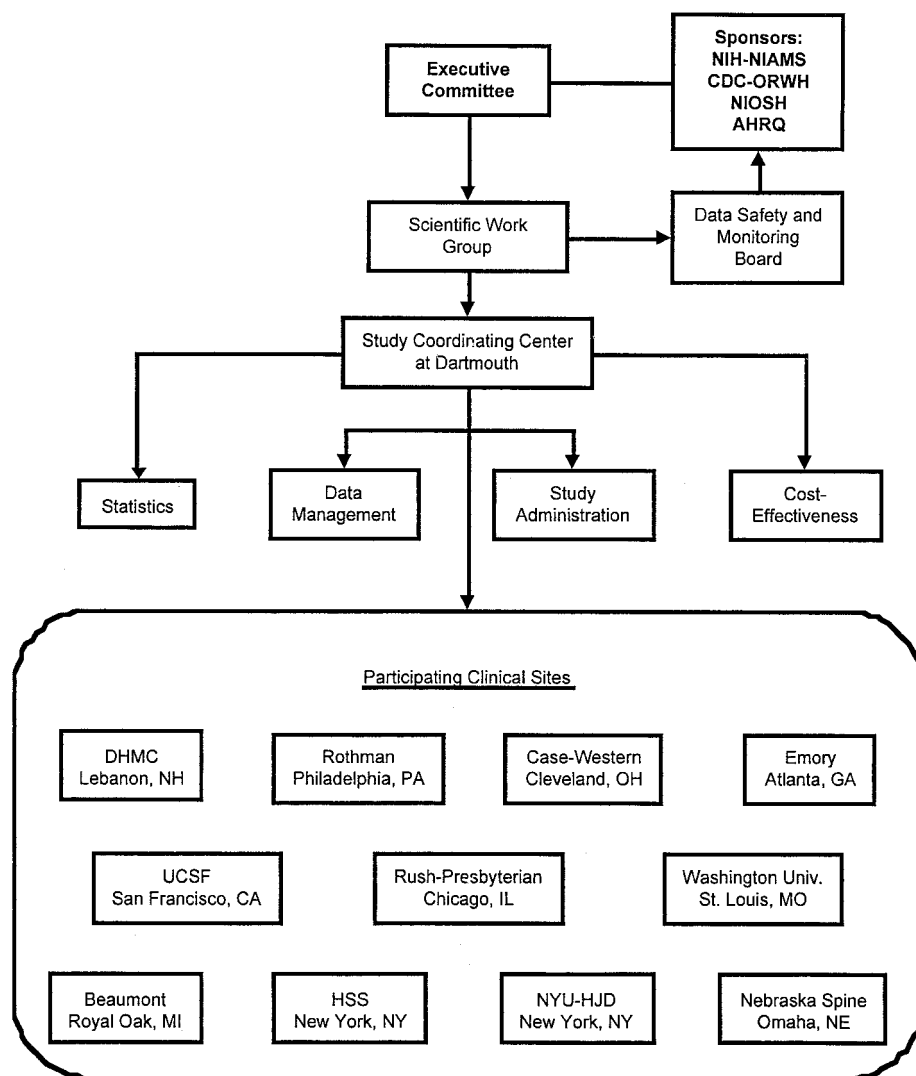


Figure 2. Organization of the SPORT project.

advantages, including the following: 1) flagging of missing, out of range, and inconsistent data items at the time of entry; 2) tracking patients to notify the nurse coordinator when follow-up visits should be scheduled and automatic generation of required paperwork; and 3) automatic generation of monthly site reports regarding enrollments into each of the randomized and observational studies.

The Data Management group also maintains a project web site that serves as a reference for protocol and policy information. Data regarding enrollments are updated daily and are also available to the sites and data coordinating center staff on the web site. All of the computers and databases are password protected, hard drives are connected to uninterrupted power supplies, and daily back-up of the individual site databases as well as the combined central database located at Dartmouth is performed to ensure data security and integrity. Data moving from participating centers to Dartmouth are encrypted and are also moved off the central server once they reach Dartmouth.

Discussion

Surgical treatments are rarely subjected to rigorous evaluation before dissemination.^{31,43,45} However, a few randomized trials have assessed the relative efficacy of surgical and nonsurgical approaches to the same problem, proving the feasibility of randomized studies involving surgery and providing some surprising results. For example, a number of commonly used surgical treatments were abandoned after randomized trials proved them ineffective.^{15,13,27,33,46} Other trials have resulted in revised indications for surgical treatment (for symptom relief rather than improved survival)³⁹ or have shown that the balance between risks and benefits is fragile for many surgical procedures.^{50,51} In other cases randomized trials proved that surgical treatments were clearly superior to the nonsurgical alternatives.^{34,37,53}

There are a number of issues that complicate the performance of randomized trials in surgery that may relate to why these types of studies are not frequently conducted. In this section we discuss some of these issues as they relate to the design and conduct of SPORT.

Equipoise and Preferences for Treatment

Randomization is not easy for patients or physicians to accept. It is rare that a patient would not have preferences regarding the alternatives where both surgical and nonsurgical options exist. It is also not surprising that physicians believe in the value of the treatments that they administer. In SPORT we decided that the best way to recruit patients into this study would be to educate them about the lack of existing scientific evidence to guide clinical decision-making for these conditions and the resulting need for randomized clinical trials. Rather than relying solely on the participating physicians to make this case, we decided to use videotaped educational materials that were developed to promote shared medical decision-making as part of our recruitment strategy. These tapes are condition-specific and provide information based on the available scientific data regarding what is and what is not known about the relative risks and benefits of surgical and nonsurgical treatments for these conditions. The programs were edited with permission from the Foundation for Informed Medical Decision Making to include an invitation to participate in SPORT.

Placebo Effects of Surgery

Placebo effects have often been discussed as a potential source of bias in surgical trials.^{5,24} In perhaps the most famous cases, patients receiving sham internal mammary artery ligation for coronary artery disease and gastric freezing for peptic ulcer in randomized, placebo-controlled trials reported improvement in their symptoms in similar proportions to the patients receiving the active treatments.^{13,41} However, both the ethics and feasibility of performing sham surgery are very much debated, and it was determined that placebo-controlled studies of spine surgery for these conditions were not possible at this time. Instead, SPORT was designed so that the “control” patients would receive an aggressive form of nonsurgical care rather than no treatment. We hope that the videotapes used in the informed consent process may help to mitigate preconceived notions regarding the efficacy of the treatment options.

Prior surgical trials have frequently relied on objective outcomes such as survival or objectively defined clinical events that are thought to be less susceptible to placebo effects. Whereas our primary outcome measures (survey instruments for measuring back and leg symptoms and health-related quality of life) are subjective, we also measure other variables such as treatment complications and work status, which are more like objective outcomes. Ultimately, the major goal of elective spine surgery is pain relief, which is necessarily a subjective phenomenon. This contrasts with trials for cardiovascular diseases, for example, where the primary endpoints are objective outcomes, such as mortality and cardiovascular events. In the end, we will not be able to determine to what extent any observed benefits of surgical treatment are attributable solely to its placebo effect, and this must be considered a limitation of SPORT.

Generalizability

In the design of clinical trials a trade-off between validity and generalizability must be struck. The results of randomized clinical trials are frequently criticized for not being generalizable because of strict institutional and patient eligibility criteria as well as bias in the selection of patients that are invited to participate.

Patient Selection. Surgical trials frequently have had restrictive eligibility criteria limiting enrollment to subjects with the lowest risks of procedure-related complications and the best prognosis.^{17,29,37} This limits generalizability to the extent that patients who are enrolled in the study differ from those for whom the surgery is regularly performed. Additionally, patients who are willing to be randomized are known to be different from those who refuse in ways that may be related to their response to treatment. Failing to adequately describe the demographic and clinical differences between the patients enrolled in the trial and those to whom the results are intended to apply is a frequent criticism of clinical trials in general.

In SPORT the results are intended to apply to patients with IDH, SpS, or DS who are surgical candidates, and the eligibility and exclusion criteria were designed to capture this population. To gather information that will allow an assessment of our study's generalizability, SPORT includes both randomized and observational trials in an effort to gather baseline and follow-up information for as many eligible patients as possible. This will allow an examination of the baseline characteristics as well as the responses to treatment for patients who agree and do not agree to be randomized.

Surgical Technique/Skill. One difference between surgery and pharmaceutical therapy is that no two surgeons do an operation exactly the same way or with exactly the same skill. Therefore, surgery cannot be standardized to the same extent that drug therapy can. In addition, surgical procedures evolve over time with the possibility that by the time any trial of a procedure is complete it will be considered obsolete. Prior surgical trials have been criticized for restricting institutional participation to high volume, academic referral centers.⁵¹ Although these concerns are frequently raised to argue against the use of randomized clinical trials to assess the efficacy of surgical procedures,³⁰ they should be of equal or even greater concern for cohort studies as well as uncontrolled case series.

In SPORT the participating physicians have agreed to general approaches to both surgical and nonsurgical therapy that represent current standard treatment for these conditions. The physicians participating in SPORT are from a range of different medical and surgical specialties (*Spine* on-line: see “Article Plus” for details), and their centers represent a wide range of institutions that perform spine surgery, including private and nonprofit, academic and nonacademic, and high volume and low volume. Any observed differences in the specific techniques and skill levels of the physicians and medical cen-

ters will be accounted for in the statistical analysis of our data. Any “advances” in treatment that occur during the conduct of the trial can be tested against the most beneficial standard approach on the completion of SPORT.

Treatment Crossover

Although rates of crossover from surgical to nonsurgical treatment have been substantial in prior studies of treatment for back conditions, the data from these studies lend some support for the performance of an intention-to-treat analysis. In the Maine Lumbar Spine Study, a prospective cohort study, 15% of the patients with herniated disc and 14% of the patients with SpS crossed over from nonsurgical to surgical treatment within the first 3 months of enrolling in the study and were treated as surgical patients in the analysis.^{3,4} The baseline characteristics of these patients were similar to patients who opted for surgical treatment initially, but their reclassification as surgical did not have any effect on the long-term outcomes by treatment group.^{1,2,26} An additional 16% of the patients with herniated disc underwent surgery between 3 and 60 months for an overall crossover rate of 30% at 5 years.¹ An additional 8% of the patients with SpS underwent surgery between 3 and 48 months for an overall crossover rate of 22% at 4 years.² The baseline characteristics and outcomes for the patients who crossed over after 3 months were not significantly different from those who stayed in the nonsurgical treatment group in either the herniated disc or SpS study.^{1,2} In Weber’s randomized trial of surgical and nonsurgical treatment for herniated disc, 26% of the patients randomized to nonsurgical treatment crossed over (all occurring within 1 year).⁴⁹ Outcomes for the patients who crossed over were not significantly different from those who stayed in the nonsurgical treatment group at 1 year (33% *vs.* 47% “good” results) or 10 years (55% *vs.* 59%).⁴⁹

Whereas our primary analyses will conform to the intention to treat principle, an as treated analysis is planned to study the potential effects of such crossovers. In these analyses the intervention group for an individual patient will depend on their actual status at the time of crossover and at each designated follow-up time. These analyses will not be based on a randomized treatment assignment and will be carefully adjusted for potential confounding factors.

Measuring Costs

In SPORT we track the use of health care services in a consistent fashion for all patients regardless of treatment arm. To achieve comparability in cost data across treatment arms, we rely on the patient as the single best source of information. Although this minimizes the potential for ascertainment bias that is inherent in relying on hospital administrative data for costing surgical (mainly hospital-based) *versus* nonsurgical (mainly non-hospital-based) care, it may result in some inaccuracies resulting from patient recall. However, as detailed ear-

lier, we are using two approaches (detailed diaries and 1-month recall window) to minimize these problems.

There are some methodologic controversies related to measuring indirect costs and benefits of employed hours and wage changes. The first is that an increase in 1 hour of work imposes some cost in terms of lost leisure time, so that simply measuring gains in earnings would tend to overstate the true utility gains to the worker. Sensitivity analysis can be used to provide bounds on how important this issue is in measuring indirect costs as well as quantifying the implications of changes in employment for government tax revenue. Second, and more importantly, is the potential for double counting the benefits of improved functioning if patients report higher levels of current health as a consequence of their being able to go back to work.²¹ However, one recent study demonstrated that financial constraints did not seem to be related to how patients responded to a survey about health functioning.³² Here again, we will use sensitivity analysis to assess the potential effects of different assumptions regarding indirect cost estimates.

A third issue is whether individual wage rates should be used to value the indirect costs and benefits of work. Using actual wages would imply that a change in 1 hour of work for an individual making \$30 per hour would be weighted five times as heavily as a change in one hour for an individual making \$6 per hour, for example. As a statistical issue, this approach could induce a great deal of “noise” into the estimates, especially if there are a few people with very high wage rates. One approach to solving the statistical noise issue is to assign wage rates according to occupation and education cells based on a larger sample from the Current Population Survey. A more philosophical issue is whether society should value the \$30 per hour individual’s change in hours of work at five times the rate of the \$6 per hour worker.⁵⁴ An alternative specification that we will test in sensitivity analyses is the calculation of indirect costs whereby each patient is assigned the average wage of the sample rather than their actual wage.

Rationale for the Observational Studies

There are at least three advantages to including observational cohorts as part of SPORT. The first is to test whether there are systematic differences in the type and severity of illness for patients who enroll in the randomized and observational study groups. As stated previously, knowing whether the randomized group is representative of the population of surgically eligible patients with low back pain is important in judging the generalizability of the study. Second, where observational cohort data can be combined with the randomized cohort data, they will increase the sample size and power of statistical analysis involving cost data and many of the secondary outcomes such as work experience and earnings. Finally, the observational group can be used in future ancillary studies and supplemental analysis of outcomes at the different surgical sites.

■ Conclusions

In other surgical specialties, large-scale randomized trials of surgical and nonsurgical treatment alternatives have been successfully conducted despite obstacles that seem much harder than those encountered in assessing the effectiveness of spine surgery. Most causes of back pain are not fatal or associated with permanent pain or disability. In addition, spine surgery is associated with less risk than most of the other types of surgery that have been studied in randomized trials. Spine surgery should not be more difficult to standardize than the other types of surgery that have been studied in this way, such as coronary bypass, carotid artery, and brain surgery. Finally, there is no reason to think that patient and surgeon preferences regarding spine surgery would be any more strongly held than those in cardiovascular surgery or (nonspine-related) neurosurgery, where large-scale randomized trials have been successfully conducted.

Acknowledgment

This study is dedicated to the memory of Brianna Weinstein.

■ Key Points

- Research addressing the causes and treatment of spinal conditions has been given high priority because of its prevalence and associated morbidity, costs, and clinical controversy.
- The Spine Patient Outcomes Research Trial (SPORT) was designed to assess the relative efficacy and cost-effectiveness of surgical and nonsurgical approaches to the treatment of common conditions associated with low back and leg pain.
- This article describes the rationale and design of the SPORT project and discusses some of its strengths and limitations.

References

1. Atlas S, Keller R, Chang Y, et al. Surgical and nonsurgical management of sciatica secondary to lumbar disc herniation: five-year outcomes from the Maine Lumbar Spine Study. *Spine* 2000;26:1179–87.
2. Atlas S, Keller R, Robson D, et al. Surgical and nonsurgical management of lumbar spinal stenosis: four-year outcomes from the Maine Lumbar Spine Study. *Spine* 2000;25:556–62.
3. Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study: II. 1-year outcomes of surgical and nonsurgical management of sciatica. *Spine* 1996;21:1777–86.
4. Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study: III. 1-year outcomes of surgical and nonsurgical management of spinal stenosis. *Spine* 1996;21:1787–95.
5. Beecher H. Surgery as placebo: a quantitative study of bias. *JAMA* 1961;176:88–93.
6. Birkmeyer J, ed. Conditions of the spine. In: *The Dartmouth Atlas of Musculoskeletal Health Care*. Chicago: AHA Press, 2000:28–58.
7. Cherkin DC, Deyo RA, Loeser JD, et al. An international comparison of back surgery rates. *Spine* 1994;19:1201–6.
8. Department of Health and Human Services. Acute Low Back Problems in Adults. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, 1994.
9. Department of Health and Human Services. Medicare Program Part B schedule for physicians services: Final Rule 405,413,415. *Health Care Financing Administration: Part III*. Rockville, MD: Department of Health and Human Services, 1991.
10. Deyo R, Weinstein J. Low back pain. *N Engl J Med* 2001;34:363–70.
11. Deyo RA. Fads in the treatment of low back pain. *N Engl J Med* 1991;325:1039–40.
12. Deyo RA. Practice variations, treatment fads, rising disability: do we need a new clinical research paradigm? *Spine* 1993;18:2153–62.
13. Dimond E, Kittle C, Crockett J. Evaluation of internal mammary artery ligation and sham procedure in angina pectoris. *Circulation* 1958;18:712–3.
14. DRG Handbook. Comparative Clinical and Financial Benchmarks. Baltimore: HCIA, Ernst & Young LLP, 2000.
15. EC/IC Bypass Study Group. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke: results of an international randomized trial. *N Engl J Med* 1985;313:1191–200.
16. EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
17. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. Endarterectomy for asymptomatic carotid artery stenosis. *JAMA* 1995;273:1421–8.
18. Fairbank J, Couper J, Davies J, et al. The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy* 1980;66:217–73.
19. Fardon D, Milette P. Nomenclature and classification of lumbar disc pathology: recommendations of the Combined Task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. *Spine* 2001;26:E93–113.
20. Friedman L, Furberg C, DeMets D. The randomization process. In: *Fundamentals of Clinical Trials*. Littleton, MA: PSG, 1985:51–69.
21. Gold M, Siegel J, Russell L, et al. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
22. Herno A, Airaksinen O, Saari T, et al. Lumbar spinal stenosis: a matched-pair study of operated and non-operated patients. *Br J Neurosurg* 1996;10:461–5.
23. Hsiao W, Braun P, Dunn D, et al. An overview of the development and refinement of the Resource Based Relative Value Scale: the foundation for reform of U.S. physician payment. *Med Care* 1992;11:1–12.
24. Johnson A. Surgery as a placebo. *Lancet* 1994;344:1140–2.
25. Johnsson K, Uden A, Rosen I. The effect of decompression on the natural course of spinal stenosis: a comparison of surgically and untreated patients. *Spine* 1991;16:615–9.
26. Keller RB, Atlas SJ, Singer DE, et al. The Maine Lumbar Spine Study: I. Background and concepts. *Spine* 1996;21:1769–76.
27. King J. Randomized trial of the Rees and Shealy methods for the treatment for low back pain. In: Morley T, ed. *Current Controversies in Neurosurgery*. Philadelphia: Saunders, 1976:89–94.
28. Laird N, Ware J. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–74.
29. Lawrie G, DeBakey M. The Coronary Artery Surgery Study. *JAMA* 1984;252:2609–11.
30. Love J. Drugs and operations: some important differences. *JAMA* 1975;232:37–8.
31. McLeod R, Wright J, Solomon M, et al. Randomized controlled trials in surgery: issues and problems. *Surgery* 1996;119:483–6.
32. Meltzer D, Weckerle C, Chang L. Do people consider financial effects in answering quality of life questions? *Med Decis Making* 1999;19:517.
33. Miao L. Gastric freezing: an example of the evaluation of medical therapy by randomized clinical trials. In: Mostelloer F, ed. *Cost, Risks, and Benefits of Therapy*. New York: Oxford University Press, 1977:198–211.
34. Moller H, Hedlund R. Surgery versus conservative management in adult isthmus spondylolisthesis: a prospective randomized study. *Spine* 2000;25:1711–5.
35. Deleted in proof.
36. Nease R, Kneeland T, O'Connor G, et al. Variation in patient utilities for outcomes of the management of chronic stable angina: implications for clinical guidelines. *JAMA* 1995;273:1185–90.
37. North American Symptomatic Carotid Endarterectomy Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N Engl J Med* 1991;325:445–53.
38. Nykyst F, Hurme M, Alaranta H, et al. Severe sciatica: a 13-year follow-up of 342 patients. *Eur Spine J* 1995;4:335–8.
39. Rahimtoola S. A perspective on the three large multicenter randomized clinical trials of coronary bypass surgery for chronic stable angina. *Circulation* 1985;72:V123–35.
40. Read J, Quinn R, Berwick D, et al. Preferences for health outcomes: comparison of assessment methods. *Med Decis Making* 1984;4:315–29.
41. Ruffin J, Grizzle J, Hightower N, et al. A co-operative double-blind evaluation of gastric “freezing” in the treatment of duodenal ulcer. *N Engl J Med* 1969;281:16–9.

42. Shvartzman L, Weingarten E, Sherry H, et al. Cost-effectiveness analysis of extended conservative therapy versus surgical intervention in the management of herniated lumbar intervertebral disc. *Spine* 1992;17:176–82.
43. Sisk J, Glied S. Innovation under federal health care reform. *Health Aff* 1994;13:82–97.
44. Sox H, Blatt M, Higgins M, et al. *Medical Decision Making*. Boston: Butterworths, 1988.
45. Spodick D. The surgical mystique and the double standard. *Am Heart J* 1973;85:579–83.
46. Thomsen J, Bretlau P, Tos M, et al. Placebo effect in surgery for Meniere's disease: a double-blind, placebo-controlled study on endolymphatic sac shunt surgery. *Arch Otolaryngol* 1981;107:271–7.
47. Torrance GW, Feeny DH, Furlong WJ, et al. Multi-attribute preference functions: Health Utility Index. *Pharmacoeconomics* 1995;9:503–20.
48. Ware J, Sherbourne D. The MOS 36-item short-form health survey. *Med Care* 1992;30:473–83.
49. Weber H. Lumbar disc herniation: a controlled, prospective study with ten years of observation. *Spine* 1983;8:131–40.
50. Wennberg D, Birkmeyer J, Lucas F. Pounds of prevention for ounces of cure: surgery as a preventive strategy. *Lancet* 1999;353:SI9–11.
51. Wennberg D, Lucas F, Birkmeyer J, et al. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. *JAMA* 1998;279:1278–81.
52. Wennberg JE, ed. *The Dartmouth Atlas of Health Care in the United States*, 1st ed. Hanover, NH: Center for the Evaluative Clinical Sciences at Dartmouth Medical School (in association with American Hospital Publishing, Inc., 1998.
53. Wiebe S, Blume W, Girvin J, et al. A randomized, controlled trial of surgery for temporal-lobe epilepsy. *N Engl J Med* 2001;345:311–8.
54. Zerbe RJ, Develly D. *Benefit-Cost Analysis in Theory, Practice*. New York: Harper Collins, 1994.

Address correspondence to

Nancy J. O. Birkmeyer, PhD
c/o SPORT/Department of Surgery
Dartmouth Medical School
One Medical Center Drive
Lebanon, NH 03756