

Outcomes Research in Orthopaedics

Robert B. Keller, MD

Abstract

A new agenda in outcomes research has developed in the past decade. The stimulus has come as the result of rapidly increasing health care costs, marked variations in utilization of health care services, and deficiencies in the research literature. Outcomes research includes methods such as analysis of large databases, small-area analysis, structured literature reviews (meta-analysis), prospective clinical trials, decision analysis, and guideline development. Clinical research should be prospective and should employ modern statistical and assessment methods. The focus of this research is on patient-oriented outcomes of care rather than on assessments of the process of care. To illustrate these applications in orthopaedics, lumbar spine fusion with internal fixation for "spinal instability" is presented as an example. Completed large-database analyses, small-area variation studies, and a meta-analysis indicate the need for clinical studies. An outline of the form and content of such a study is presented.

J Am Acad Orthop Surg 1993;1:122-129

During the past 5 to 10 years a new term has appeared in the medical vocabulary—"outcomes research." The purpose of this article is to define and describe this new concept, particularly as it relates to orthopaedic surgery. Additionally, by using a clinical example, the methods underlying this concept will be clarified.

What is outcomes research and why do we need to be concerned about it? Outcomes research can be simply defined as refined and enhanced clinical research. In this research there is an important focus on patient-based outcomes as opposed to measures of process of care. Patient-based outcomes are assessments that measure the results of care as they are perceived by patients. They include factors like pain, function, satisfaction, and quality of life. Process measures include such factors as radiographic appearance, range of motion, and laboratory results.

Additional new methodologies, such as large-database analysis,

small-area analysis, meta-analysis, and decision analysis, have become an important part of outcomes assessment, but clinical research remains the basis of the concept.

Factors in Rethinking Clinical Research Methods

At the outset, we need to understand the factors that have been the stimulus for this major rethinking of the clinical research methods we understand and have relied on for so long. Several important factors have developed in the past 15 or so years. Singly and together, they make it clear that we who practice medicine need to rethink our current knowledge base and how we develop new information.

The Rising Costs of Health Care

It seems clear that the dramatic increase in health care costs over the past 30 years is the major factor

underlying the outcomes agenda. The percentage of gross domestic product spent on health care in the United States has risen from 5.2% in 1960 to 14.4% in 1992—the highest percentage among the industrialized nations. In 1990 (the most recent year for which comparable data are available), the United States spent 12.2% of its gross domestic product on health care, compared with 8.5% for Canada, 6.3% for Japan, and 8% for Germany.¹ Various broad measures of health status, such as life expectancy and infant mortality, indicate that our extra expenditures produce no obvious benefit. That is not to say that the increased expenditures in the United States do not produce higher quality or more effective care. The problem is that we have no information to prove the point.

Practice-Pattern Variations

In 1973 Wennberg and Gittelsohn² published their first article on the subject of variations in practice and utilization patterns in medical care, which provided a major stimulus to more rigorous evaluation of clinical practice.

Dr. Keller is Executive Director, Maine Medical Assessment Foundation, Augusta, Me; Adjunct Professor of Surgery and Community and Family Medicine, Dartmouth Medical School, Hanover, NH; and Associate Professor of Orthopedic Surgery, University of Massachusetts Medical School, Worcester.

Reprint requests: Dr. Keller, Maine Medical Assessment Foundation, Box 4682, 18 Spruce Street, Augusta, ME 04330.

Epidemiologists have typically expressed the incidence of disease in terms of the rate of occurrence of a condition (the number of episodes per 100,000 population). Wennberg and Gittelsohn applied similar methods to study the utilization or consumption of health care services. They further refined the method by developing "small areas," geographic regions surrounding hospitals at which the majority of local residents receive care. It turned out, contrary to what one might think, that there are marked differences in hospital admission and surgical rates between small areas within states. As one looks more widely, there are also significant differences between states, regions, and nations. It is also important to note that all health care systems, regardless of their organization or financing design, demonstrate this kind of variation.

Within orthopaedics, there are few conditions that do not show variations. Hip fracture and multiple trauma are examples of low-variation conditions. Essentially every other condition or procedure in the specialty shows striking variations in hospital and surgical use rates.³

The conclusion reached by those who have carried out these studies is that after careful statistical adjustments for factors such as age and sex, the wide variations that exist are not appropriate. If the high rate of utilization represents the "right rate," then those below that level are being underserved. If the low rate is correct, then those above it are receiving excessive care. The problem is that we do not know what the so-called right rate is, but it does seem clear that all the rates cannot be correct. Outcomes research hopes to answer this conundrum.

Deficiencies in the Clinical Literature

The major source of information for clinicians is the published litera-

ture. Directly or indirectly, almost all knowledge in orthopaedics is based on information that has appeared in journals and texts. Researchers and investigators write them, teachers teach from them, students read them, board examinations are based on them, and those in practice rely on them in their daily practice of orthopaedics. Without the core of information based in the scientific literature, we would practice folk medicine.

In recent years that fundamental basis of knowledge and learning has come into question. The questions come from two sources. First, authors have critically reviewed certain areas of the clinical literature regarding its quality and accuracy. They have found significant problems. Gartland⁴ and Gross⁵ have both analyzed the literature of hip arthroplasty. Each found significant flaws in it. Faulty research design, erroneous statistical analysis, and a lack of focus on patient-oriented outcomes of treatment were noted.

The second source comes from a new technique of scientific literature review known as meta-analysis. In this method, data from many articles are pooled to form a larger mass of information for statistical analysis. Ideally, only randomized trials qualify for meta-analysis, but few of these have occurred in orthopaedics. With care, one can broaden the criteria to include other reports. Meta-analyses have been published for several orthopaedic conditions, including hip fracture, lateral epicondylitis, and lumbar spine fusion.⁶⁻⁹

The consistent finding in these reports has been the lack of randomized trials, inadequate study design, lack of standardized definitions and measures, poor descriptions of patients, inadequate and unclear follow-up, and little or no evaluation of patient-oriented outcomes of care. Indeed, some attempts at meta-

analysis have not been possible because the available literature is so weak.^{7,9}

If the literature on which we so heavily depend has such significant deficiencies, it is perhaps not surprising that practice-pattern variations exist. There is simply not a firm knowledge and research base on which the clinician can rely in clinical decision making.

Outcomes Research Methodologies

Outcomes research in its broadest context involves a number of different methods—literature review, large-database analysis, small-area analysis, prospective clinical trials, decision analysis, and development of clinical guidelines. In the large, federally funded Patient Outcomes Research Teams studies,¹⁰ essentially all of these methods are utilized. However, these techniques may be used independently. For instance, meta-analysis is one method within outcomes research, but this kind of analysis is often undertaken as an independent effort.

Literature Review

An important step in all research is the need to review what is known about a subject up to the current time. Ideally, one would carry out a meta-analysis of the literature for each and every project.¹¹ The object of meta-analysis is to gather comparable data from a number of different sources and combine those data to create a larger and more statistically significant pool of information for analysis. In each analysis, strict rules for inclusion and exclusion of data from different sources must be developed. Reader bias in selection and interpretation of articles is thus avoided. Because meta-analysis is time consuming and expensive (\$30,000 to \$50,000 per analysis is not

unusual), and the literature may be so deficient as to defy a high-quality meta-analysis, this step may not be necessary or useful. A “structured literature review” in which one applies many of the rules of meta-analysis may suffice. The more typical “narrative review,” in which an author picks and chooses which articles to quote and emphasize, is subject to significant bias.

Large-Database Analysis

This method utilizes analyses of large databases, such as the Medicare files. It should be noted that these are primarily claims data, which may be subject to significant error and may require great skill to interpret. Other claims databases and state-level hospital discharge data abstracts can also be useful. From these sources one can carry out epidemiologic studies and limited outcomes analyses on factors such as mortality, length of stay, complications, and reoperations.¹²

None of these databases is perfect, and in carrying out analyses and drawing conclusions, analysts must be experienced and must exercise caution. However, there is a tremendous amount of valuable information in them.

Small-Area Analysis

This form of analysis is a methodologic subset of large-database analysis, in that one needs to access a large database to carry it out. Small-area analysis is of specific interest because it demonstrates to physicians (and others) that there are significant inconsistencies in their practice patterns. It serves the important and useful purpose of engaging practitioners in the process of analysis, feedback, research, and change in practice patterns.

Prospective Clinical Trials

Clinical research should be conducted prospectively, ideally through

randomized clinical trials. Recognizing that it is not always possible to randomize patients for many kinds of medical and surgical treatments, there are several other study designs¹³ that can reasonably effectively control for various biases. Retrospective studies should be avoided. It is extremely difficult to recover valid and accurate outcomes information from records that were not set up for the purpose of a specific study. Numerous methodologic problems can occur.

It is most important to carefully plan the study so that the hypotheses one wishes to test will, in fact, be tested. This implies that the investigators will design proper data-collection instruments, calculate adequate sample size, plan careful follow-up protocols for all patients, collect information relevant to patient-oriented outcomes of care, and conduct proper statistical analyses. Patient outcomes assessment includes categories such as satisfaction, function, pain, utility, and quality of life. Evaluation instruments are available to accurately measure many general health factors.¹⁴ It remains for the specialties to develop standardized and valid instruments for the conditions they treat.

Ideally, outcomes studies should involve alternative forms of treatment (e.g., a comparison of surgery and medical treatment for a given condition). Case-series reports (the most common in the literature) provide very biased information because one never knows how patients might have fared with another treatment, or perhaps no treatment at all.

Decision Analysis

This is a relatively new concept adapted from the business world. The statistical results of clinical research can be translated into a series of probabilities and placed

into an algorithm or decision tree, enabling one to numerically estimate the likelihood of various treatment outcomes based on patients' health states, complications, and specific outcomes. Outcomes can be weighted according to their desirability (e.g., from perfect health to death). Combining the probabilities and the values assigned to various outcomes can help to determine the optional strategies that are most likely to maximize good results.

The analysis may also point out where critical information is missing (and research is needed) or which decisions are most critical in influencing clinical results. Decision analysis provides a numerical probability of a given outcome.

Clinical Guidelines

Guidelines are an important product that can be developed from outcomes research. One of the major problems in developing valid and useful guidelines is the fact that accurate information and data to inform the guideline process have not been available. Thus, the deficiencies of clinical research also restrain the development of guidelines. As the results of improved clinical outcomes research become available, they can be used to develop high-quality practice guidelines.

A Clinical Case Example: Lumbar Spine Fusion

To demonstrate the components and methods involved in outcomes research, it would be helpful to use a specific clinical example. I have chosen instrumented lumbar spine fusion because it represents a new technology, shows wide variation in utilization, and is a controversial procedure.

With the development of several spinal fixation devices in the past

decade, there has been rapid growth in the rates of lumbar fusion. The increased utilization of this procedure has outpaced the population growth or any known risk factors that might produce increased patient need or demand for the procedure.¹⁵ It would appear that the increase in utilization of the procedure has been driven in part by the availability of a new technology. The question remains: Has the availability of this new technology improved patient outcomes in a way that can justify the increase in utilization?

This clinical situation appears ideally suited for outcomes research. In fact, although additional clinical research is required to establish more precisely the role of this procedure in patient care, several steps in the research process have already been undertaken.

Large-Database Analysis

Analyses of spine fusion rates across large national regional areas have been performed. As with all elective surgical procedures, significant variations are seen. Rates of spine fusion across the four major regions of the nation have been determined for the years 1988 to 1990 through analyses of the National Center for Health Statistics database.¹⁵ They indicate a 56% greater likelihood of spine fusion for Midwesterners than for residents of the Northeast (Taylor V, Deyo R: written communication, August 1993) (Fig. 1). An analysis of 1990 fusion rates among residents of the five largest counties in the state of Washington reveals a variation of 240% (Fig. 2). In another analysis of the Washington database, Deyo et al¹⁶ determined that the rate of in-hospital complications for disk excision procedures was 5.4%, which increased to 12.1% when fusion was combined with disectomy.¹⁷

The only way that these variable rates would be reasonable is if there

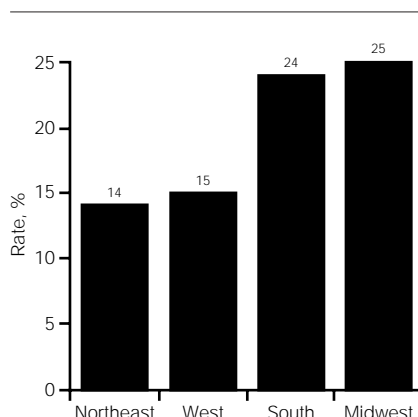


Fig. 1 Average annual rates of performance of lumbar fusion per 100,000 adults (age- and sex-adjusted to the 1990 US population) for the four large geographic regions of the United States in the period 1988 to 1990. Fusion was performed 180% more frequently in the Midwest than in the Northeast.

were major differences in underlying spine pathology, work conditions, or injury rates, but these differences were not identified. The more likely explanation relates to the differing practice styles of orthopaedists (and, more recently, neurosurgeons) as they reflect their beliefs about the efficacy and effectiveness of this procedure. These data do not indicate that any one of these regional or state rates is prefer-

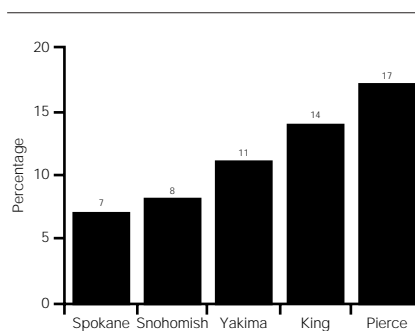


Fig. 2 Percentage of patients in the five most highly populated counties in the state of Washington who underwent lumbar spine procedures in 1990 who also underwent fusion (unpublished data provided by Victoria M. Taylor, MD, Seattle).

able. Indeed, none of them may be the so-called right rate. But it would be difficult to defend all of them as being appropriate. There is a strong implication that there may be overuse of the procedure in the Midwest or underutilization in the Northeast, or perhaps both.

Small-Area Analysis

We have studied the utilization of lumbar fusion across the 72 hospital-service areas of Maine, New Hampshire, and Vermont. Each of these areas contains at least one hospital and has one or more orthopaedic surgeons or neurosurgeons practicing within it. While lumbar disk excision and cervical procedures vary only minimally among the three states, the rate of lumbar fusion across the region varies by a factor of 3.6 (Fig. 3). Two clusters of service areas in the three states have significantly higher ($P < .01$) utilization rates than the rest of the region.

A study group of orthopaedic surgeons from the three states has evaluated these data and cannot explain the variations on the basis of population, injury, disease, or other demographic factors. The only obvious variable is the presence of spine-fellowship-trained surgeons in those service areas where the rates are high. Subspecialty orthopaedists are located only in the areas with the highest fusion rates, with two exceptions. There are fellowship-trained surgeons in the two academic medical centers located in Vermont and New Hampshire. Because of the wide referral areas of these centers, it is possible for spine surgeons to be busy in their subspecialty without high per capita rates of surgery for the populations they treat. However, this factor is often not the case for community-based surgeons.

Fellowship-trained spine surgeons have the greatest expertise in this procedure, and one would properly expect them to perform most of

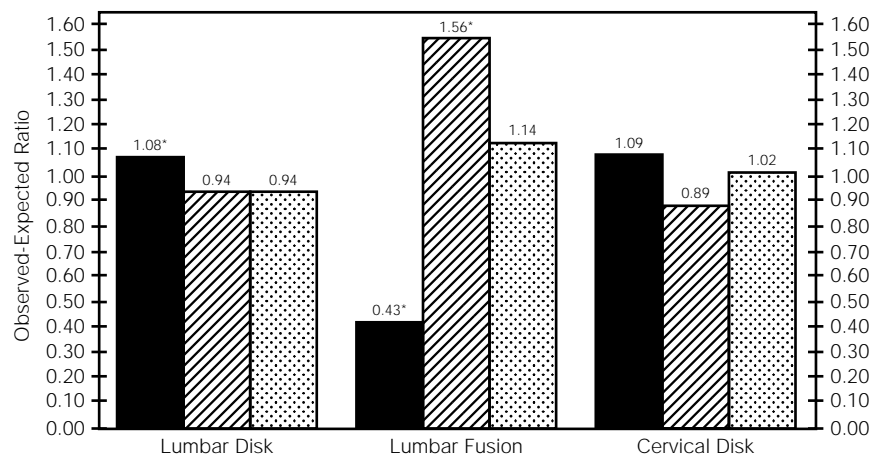


Fig. 3 Ratios of observed to expected rates for three commonly performed spine procedures for Maine (solid bars), New Hampshire (hatched bars), and Vermont (dotted bars) (* = $P < .01$). If the surgical practice patterns in the three states were similar, the ratios would be 1.00, indicating no variation among the states. There are only minor differences in the utilization of lumbar discectomy and cervical disk surgery; however, there is a 360% greater utilization of fusion procedures in New Hampshire compared with Maine. (Adapted with permission from Taylor VM, Deyo RA, Cherkin DC, et al: Low back pain hospitalization: Recent U.S. trends and regional variations. *Spine* [in press].)

the fusion surgery in their hospitals. If the provision of this service were consistent, and patients from areas outside the practice locations of the spine specialists were referred to them, one would anticipate that surgical rates across the region might be fairly level. That is because the utilization of a service is counted back to the area of residence of the patient. What our data demonstrate is that fusion rates vary according to where the experts are in practice. Patients who reside in service areas where spine surgeons are in practice have a much greater likelihood of undergoing a fusion than those who reside in adjacent service areas.

It should be apparent that one cannot draw conclusions from evaluating the volume of surgery performed by an individual practitioner. Only when population-based rates are determined can the rate of utilization of the procedure be calculated. A given surgeon could perform a large number of operations but provide a low rate of those services to the pop-

ulation being served (as in the academic centers). The converse is also true. A surgeon doing a small or moderate number of procedures might have a high per capita rate because the population served is small (as in the northern New England service areas noted).

We are left with the same question raised by the large-database analyses: which of these rates is the right rate? With small-area analyses, however, the questions of appropriateness of treatment are even more compelling. Why do residents of one service area have over three times the likelihood of undergoing a lumbar fusion as those in a community 20 miles away? Until these analyses are undertaken and presented to physicians, they have no idea that the variations exist and how their practice patterns compare with those of their colleagues in the region.

Literature Review

As with all investigations, consideration of outcomes research to

study lumbar fusion should be based on what the literature can tell us about the procedure. A meta-analysis of this literature has been published.⁸ The authors used standard meta-analytic techniques in their review. Their conclusions and recommendations are similar to those in other published meta-analyses. Their analysis of the available literature pertaining to lumbar fusion revealed that there were no randomized clinical trials of the procedure. They found an average of 68% satisfactory results (range, 16% to 95%), a pseudarthrosis rate of 14%, and a rate of painful donor graft sites of 9%. The study also indicated similar clinical success rates for instrumented as opposed to noninstrumented fusions.

The conclusion of this review is that better research is urgently needed on both the effectiveness (does the technology work when broadly applied at the community level?) and the appropriateness (is the technique being utilized for the proper patients?) of lumbar fusion. As should now be clear, the same questions can be asked about most orthopaedic procedures.

Designing a Prospective Clinical Study

To frame a prospective study, one must first develop a hypothesis. It would be difficult in one study to evaluate all aspects of lumbar fusion. However, it is always desirable to evaluate alternative methods of treatment. One might wish to study a condition for which fusion may be a treatment option. Fusion or nonsurgical treatment for "spinal instability" is an example. One could also evaluate different fixation devices applied to similar cohorts of patients to learn whether some are preferable to others. There are numerous hypotheses that can be generated.

One can make the case that these kinds of studies should have been conducted prior to the wide dissemination of spinal instrumentation technology, but this pattern of broad dissemination and utilization of new technologies is very common, and the questions still need to be answered.

It is important to emphasize at this point that outcomes research is a team effort. One of the reasons for the deficiency of the current literature is that many research efforts have been carried out without the benefit of a team approach. One will need the support of a research methodologist, a biostatistician, and perhaps a survey methodologist and an epidemiologist. A recent review of methodologies and statistical methods in the spine literature noted statistical deficiencies in 54% of studies and questionable conclusions based on misleading significance testing in 46%.¹⁷ The need for expert support in these disciplines is clear. For more complex studies, colleagues such as health economists, sociologists, and others may be required. Clinicians are critical to the research, but they cannot design and carry out these studies alone.

Assume that we wish to study the outcomes of spine fusion for spinal instability, and we wish to compare patients who undergo fusion for this diagnosis with a group who are treated nonoperatively. The first step is to find out how many patients are required in each treatment arm. That will require the assistance of a research methodologist who can calculate the number of patients required to measure meaningful differences in outcome—an exercise known as power analysis.

Next, one must decide how to select the patients for each treatment group. If physicians and patients were completely uncertain about which treatment is better, it would

be possible to randomize patients into different treatment groups. As an example, Herkowitz and Kurz¹⁸ have carried out a prospective study of patients with spinal instability secondary to degenerative spondylolisthesis. Alternating patients were assigned prospectively to an instrumented fusion group or a nonfusion group. While alternating is not a pure form of randomization, this study does demonstrate the important principle of prospectively evaluating patients undergoing different treatments. True randomization is difficult in most clinical situations because physicians and patients may have distinct preferences for a specific treatment and might therefore be uncomfortable with randomization.

An alternative method would be to randomize the physician rather than the patient.¹³ In that situation patients would be randomly assigned to surgeons, who would apply the treatments they prefer.

Another design is the cohort study. In this concept, patients and physicians arrive at treatment decisions in the usual way. At that point, patients are enrolled in a prospective protocol. In the case of spinal instability, patients who elect to undergo spine fusion are enrolled in the surgical cohort and those being treated nonoperatively are entered in the other cohort. Data are collected prospectively from both groups.

By carefully collecting patient-specific information in a cohort study, it may be possible to stratify reasonable comparison groups to contrast the outcomes of the different procedures. In some situations, the two groups may be sufficiently different in their presenting conditions that comparisons become impossible. If appropriate data are carefully collected, analysts will be able to make this important determination and indicate which set of analyses is possible.

Of greatest importance is determining the kinds of information to collect. Some of this may be obvious, but much is not. One of the great deficiencies in current publications is that the correct information is not solicited from patients at the time of the study. Clinicians generally know what they would like to learn from patients, but they frequently do not have the skill to frame questions in order to get the information they seek. In addition, clinicians may not know what is really important to patients about the results of their care. Survey methodologists play an important role in developing and testing patient questionnaires. They may need to interview focus groups of patients who have the condition or who have undergone spine surgery in order to learn what their concerns are.

There should be an emphasis on patient-oriented outcomes of care. For example, patients are not particularly interested in whether they have a solid spine fusion, but they are interested in factors such as pain, function, and quality of life. The degree of satisfaction and quality of life is more relevant to patients than is range of motion or radiographic evidence of fusion. Certainly, there is ample evidence that good clinical outcomes can occur despite failed fusion, and vice versa. Process measures such as strength and range of motion may not be related to outcomes measures. Often, both process and outcome need to be evaluated.

An additional problem is that there are few, if any, standardized definitions and measurements that all investigators have agreed to use. Thus, even if an article contains valid information, it is difficult to compare with others. For example, there is no broadly accepted definition of spinal instability. There are various radiographic criteria,¹⁹⁻²² which are felt to be of variable validity. Others advocate intraoperative measurements²³ or physical mea-

tures.²⁴ The point is that none of these measures has been broadly accepted and validated. If one cannot define the condition being studied, the research effort becomes most difficult to undertake.

One of the urgent needs in outcomes research is the creation of high-quality, standardized, broadly accepted, validated survey instruments. This single step would improve the quality of all reports and make possible meaningful comparisons of various treatments and conditions. In part, this issue is being addressed in the field of low back pain and lumbar spine surgery. The North American Spine Society has supported the development of a patient-oriented outcomes questionnaire. Its broad adoption and use across many clinical investigations will provide a common set of outcomes information. Thus, investigators will shortly have available at least some of the instruments they need to evaluate the outcomes of lumbar surgery in a consistent manner, but much work remains to be done.

Even with adoption of standardized measures, additional data will be required by specific outcomes projects. Those comparing the outcomes of fusion and nonoperative treatment for instability will need to collect very specific information (e.g., fusion rates, implant failures, surgical and medical complications, reoperations, and drug reactions) that might not be part of another study. The important thing is to utilize tested and broadly accepted instruments whenever they are available and to obtain expert assistance in designing and implementing new measures when necessary.

Prospective collection of data is essential. Only in this manner can the investigator be sure that all essential information is collected, that patients are appropriately categorized, and that data are collected at consistent time intervals for every patient. It is

very important to attempt to follow up all patients. If a number of patients are lost to follow-up, it is very difficult to draw proper conclusions. For instance, if a large number of patients with excellent results from spine fusion fail to return for follow-up, the results will be biased in favor of those who do poorly.

One of the problems in analyzing the outcomes of spine surgery is that long follow-up is necessary. While information can be reported at various intervals, one must attempt to carry out long-term studies.

Expert assistance is required in performing data analyses in outcomes projects.¹⁸ Relatively few clinicians have the expertise to independently conduct the various analyses and statistical significance testing. Careful statistical analysis is a critical step. Given modern statistical techniques, it may be possible to carry out manipulations such as multiple regression analysis and obtain statistically significant but clinically meaningless information. Conversely, clinically important differences might be overlooked if statistical significance is lacking; high-quality statistical analysis might be more revealing.

Finally, when information is reported, the research methods, patient-group selection process, and analyses utilized in the study must be clearly stated so that readers can clearly understand and extract the material, and perhaps even attempt replication of the results. Common definitions and standardized reporting methods will permit comparison of different techniques and methodologies and aggregation of data across reports.

Conclusions

In considering outcomes research as applied to spinal instability, we have been able to describe many of the

methodologies of this discipline as they might apply to a specific orthopaedic condition and surgical procedure. In formulating a research approach to this clinical entity, two aspects have become clear. First, we can see that outcomes research is not markedly different from clinical research as we know it. The differences relate primarily to improved research methodologies and a focus on patient-oriented outcomes of care. Second, in considering research on fusion for spinal instability, we find that there are major hurdles to overcome before one can even begin such an effort. At the outset, there is no agreement on how to define and measure the condition referred to as "spinal instability."

The issues discussed in this article put policy makers, patients, and payers in a position to make a powerful argument: "Demonstrate to us that this highly variable, very expensive, and complicated surgery for spinal instability is cost effective and really makes patients better. If you cannot, we will no longer pay for it." At present, we cannot agree on what spinal instability is, and there are no accurate data about patient outcomes. How can we presume to know who should undergo this procedure and justify to payers and patients the significant expenditures, complications, and uncertain outcomes associated with this kind of major surgery?

It thus seems imperative to perform careful studies and analyses to determine whether the entity that appears to demonstrate radiographic or imaging evidence of instability is in fact correlated with a measurable clinical presentation of pain, other symptoms, and disability. Having accomplished that task, one must then proceed to assess whether lumbar fusion produces a better outcome for patients than might result from other treatment approaches.

Finally, it should be clear that carrying out outcomes research is not an easy task, but it should also be evident that there are no real alternatives to conducting this kind of investigation. The urgent challenge

is for orthopaedic surgeons to become involved in these initiatives.

Acknowledgments: The author gratefully acknowledges the advice and assistance of Richard A. Deyo, MD, MPH, and Victoria M.

Taylor, MD, MPH, in reviewing this manuscript and in providing spine surgery data.

Supported by grant No. HS 06344 (The Back Pain Outcome Assessment Team) and grant No. HS 06813 (Outcomes Dissemination: The Maine Study Group Model) from the Agency for Health Care Policy and Research.

References

1. *Health Care Resource Book*. Washington, DC: House Committee on Ways and Means, 1993. US Government Printing Office, publication WMCP:103-4.
2. Wennberg J, Gittelsohn A: Small area variations in health care delivery. *Science* 1973;182:1102-1108.
3. Keller R, Soule DN, Wennberg JE, et al: Dealing with geographic variations in the use of hospitals: The experience of the Maine Medical Assessment Foundation Orthopaedic Study Group. *J Bone Joint Surg Am* 1990;72:1286-1293.
4. Gartland JJ: Orthopaedic clinical research: Deficiencies in experimental design and determinations of outcome. *J Bone Joint Surg Am* 1988;70:1357-1364.
5. Gross M: A critique of the methodologies used in clinical studies of hip-joint arthroplasty published in the English-language orthopaedic literature. *J Bone Joint Surg Am* 1988;70:1364-1371.
6. Lu-Yao GL, Keller RB, Littenberg B, et al: Outcomes after displaced femoral neck fractures: A meta-analysis of 106 published reports. *J Bone Joint Surg Am* (in press).
7. Labelle H, Guibert R, Joncas J, et al: Lack of scientific evidence for the treatment of lateral epicondylitis of the elbow: An attempted meta-analysis. *J Bone Joint Surg Br* 1992;74:646-651.
8. Turner JA, Ersek M, Herron L, et al: Patient outcomes after lumbar spinal fusions. *JAMA* 1992;268:907-911.
9. Turner JA, Ersek M, Herron L, et al: Surgery for lumbar spinal stenosis: Attempted meta-analysis of the literature. *Spine* 1992;17:1-8.
10. AHCPR Program Note: Medical treatment effectiveness research. Rockville, Md: Agency for Health Care Policy and Research, US Dept of Health, Education, and Welfare, March 1990.
11. L'Abbé KA, Detsky AS, O'Rourke K: Meta-analysis in clinical research. *Ann Intern Med* 1987;107:224-233.
12. Wennberg JE, Roos N, Sola L, et al: Use of claims data systems to evaluate health care outcomes: Mortality and reoperation following prostatectomy. *JAMA* 1987;257:933-936.
13. Rudicel S, Esdaile J: The randomized clinical trial in orthopaedics: Obligation or option? *J Bone Joint Surg Am* 1985; 67:1284-1293.
14. Liang MH, Fossel AH, Larson MG: Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990;28:632-642.
15. Taylor VM, Deyo RA, Cherkin DC, et al: Low back pain hospitalization: Recent U.S. trends and regional variations. *Spine* (in press).
16. Deyo RA, Cherkin DC, Loeser JD, et al: Morbidity and mortality in association with operations on the lumbar spine: The influence of age, diagnosis, and procedure. *J Bone Joint Surg Am* 1992;74: 536-543.
17. Vrbos LA, Lorenz MA, Peabody EH, et al: Clinical methodologies and incidence of appropriate statistical testing in orthopaedic spine literature: Are statistics misleading? *Spine* 1993;18: 1021-1029.
18. Herkowitz HN, Kurz LT: Degenerative lumbar spondylolisthesis with spinal stenosis: A prospective study comparing decompression with decompression and intertransverse process arthrodesis. *J Bone Joint Surg Am* 1991;73:802-808.
19. Dupuis PR, Yong-Hing K, Cassidy JD, et al: Radiologic diagnosis of degenerative lumbar spinal instability. *Spine* 1985;10:262-276.
20. Friberg O: Lumbar instability: A dynamic approach by traction-compression radiography. *Spine* 1987;12:119-129.
21. Stokes AF, Frymoyer JW: Segmental motion and instability. *Spine* 1987;12: 688-691.
22. Dvořák J, Panjabi MM, Novotny JE, et al: Clinical validation of functional flexion-extension roentgenograms of the lumbar spine. *Spine* 1991;16:943-950.
23. Ebara S, Harada T, Hosono N, et al: Intraoperative measurement of lumbar spinal instability. *Spine* 1992;17:S44-S50.
24. Paris SV: Physical signs of instability. *Spine* 1985;10:277-279.