

Using Outcomes to Improve Quality of Research and Quality of Care

Richard A. Deyo, MD, MPH

Your clinic has decided to undertake a quality-improvement initiative for patients with low back pain. A steering committee has recommended implementing a set of evidence-based guidelines developed through the sponsorship of the Agency for Health Care Policy and Research. Strategies are now in place to modify your guidelines and your colleagues' prescribing practices with regard to bed rest, use of imaging studies, and days off from work. Always the skeptic, you ask your medical director, "How will we know if following these guidelines has done any good?"

For an increasing number of clinicians and health care administrators, the answer to this question would be that we should examine patient outcomes. But what does outcomes mean, and why has *outcomes research* become a buzzword of the 1990s?

Several pressures have led to interest in studying outcomes of care and to the growing importance of evidence-based medicine in general. Perhaps the most obvious has been the rapid increase in health care costs, which leads health care purchasers, employers, the public, and the government to ask, "Are we getting our money's worth?" This question seems especially relevant because, despite higher costs, many public health statistics regarding morbidity and mortality are worse in the United States than in many other developed countries.¹ Second, accumulating evidence shows enormous geographic variations in the style of medical care² and even specialty variations in the care of patients with apparently similar conditions.³ Such wide variations in practice style have suggested to many observers that some clinical practices and their resultant costs are idiosyncratic, based on variations in training, local habits, and differing opinions rather than on firm evidence of what produces the best patient results. The implication is that some care might be unnecessary. The questions from those outside the medical profes-

sion are, "If you guys know what you are doing, why is there such wide variation in care? And why should we pay for care that exceeds some norm or average?" Such questions constitute a call for accountability by health care providers: justify the high costs in terms of good outcomes.

What Are Outcomes?

Traditionally physicians have thought of outcomes in terms of physiologic phenomena. For example, suppression of cardiac arrhythmias would be seen as an appropriate outcome for antiarrhythmic drug therapy in a patient who has ischemic heart disease. A decrease in the erythrocyte sedimentation rate might be seen as an outcome for patients with rheumatoid arthritis who are receiving disease-remitting therapy. Reduction in prostate size might be seen as the appropriate goal of drug therapy for benign prostatic hypertrophy. Ultimately, however, these physiologic and anatomic end points are termed *surrogate outcomes*. We presume that they are closely linked with the ultimate outcomes of greatest interest to patients and to society, such as symptom relief, the ability to perform normal daily activities, and survival. There are, however, many sobering examples of surrogate outcomes that were poor markers for these ultimate outcomes of interest. Some examples are shown in Table 1.⁴⁻⁸

All of these examples illustrate a hazard of depending on surrogate outcomes to judge the effects of therapy and support the argument that if we regard symptom relief, daily functioning, and survival as the major goals of therapy, then we must measure them directly rather than inferring them from physiologic changes. Unfortunately, clinical

Submitted, revised, 10 Sept 1998.

From the Center for Cost and Outcomes Research, Department of Medicine, and Department of Health Services, University of Washington, Seattle, Washington. Reprints will not be available.

Supported in part by grant No. HS-08 194 from the Agency for Health Care Policy and Research.

Table 1. Examples of Dissociations Between Surrogate Outcomes and End Results.

Treatment	Surrogate Outcomes	End Results
Encainide, flecainide for ventricular arrhythmia after myocardial infarction ⁴	90% suppression of complex ventricular ectopy	Mortality twice as high as with placebo
Clofibrate for hypercholesterolemia ⁵	Lower cholesterol levels, fewer ischemic heart disease events	Mortality 25% greater than with placebo
Plasmapheresis for rheumatoid arthritis ⁶	Lower erythrocyte sedimentation rate, complement levels	No improvement in pain relief, function, or number of inflamed joints
Finasteride for benign prostatic hyperplasia ⁷	Shrinkage of prostate size	No improvement in urinary frequency or urgency
Biofeedback for low back pains ⁸	Reduced electromyographic activity in paraspinous muscles	No significant reduction in pain

research has historically focused on just such physiologic outcomes, and only recently have investigators begun to incorporate measures of symptoms and function into many clinical trials.

These considerations in part explain the recent growth of questionnaire measures for assessing patient symptoms, function, and health-related quality of life. The latter term has been adopted because of the recognition that quality of life depends on many things beyond the control of medical care, including income, safe housing, job opportunities, and many other features of the social, political, and economic environment. The current generation of questionnaires for measuring health status or health-related quality of life fuse social science methodology and clinical expertise to quantify important but subjective phenomena.

A common complaint about questionnaire data is that they are soft as opposed to the harder outcomes of physiologic measures. The boundary, however, between hard and soft data is often indistinct. We might judge the hardness of data by their objectivity (physician report versus patient observation), preservability (radiographic or histologic specimen), or dimensionality (eg, a hematocrit measurement versus the observation of paleness). On the other hand, some methodologists argue that reproducibility of a result is the crucial attribute of hardness, and by this yardstick many questionnaire measures are at least as hard as widely accepted clinical measures.⁹ Table 2 illustrates some examples in which the reproducibility of questionnaire measures actually exceeded the reproducibility of expert clinical judgments.¹⁰⁻¹⁵

Reproducibility aside, it is possible to show that questionnaire results correlate with other health phenomena of obvious importance. For example, Figure 1 shows survival curves among middle-

aged men in a national survey who responded at baseline to a question regarding their overall health.¹⁶ The 10-year mortality was strongly associated with the respondent's own judgment about whether his health was excellent, very good, good, fair, or poor. Furthermore, the survival differences were substantial, ranging from about 60 percent to 95 percent. Although little is known about how the respondents made these health judgments at baseline, this simple subjective report obviously had substantial prognostic power. Associations of this sort provide evidence for the validity of many widely used health status questionnaires.

Outcomes Management

Most observers make a distinction between outcomes management and outcomes research. Outcomes management refers to the use of outcome measures in the course of routine clinical care. For

Table 2. Studies of the Reproducibility of Patient Reports and Expert Evaluations.

Subjective Reports	Kappa*	Expert Evaluation	Kappa*
Cough	0.87	Abnormal findings, throat examination	0.37
Runny nose	0.75	Abnormal tympanic membranes	0.42
Health history questionnaire	0.79	Dorsalis pedis pulse, present or not	0.51
Function: sickness impact profile	0.87	Ankle reflexes normal	0.50
Pain: visual analog scale	0.94	Radiologist agreement whether lumbar spine radiographs are normal	0.51

From Deyo et al,¹⁰ Wood et al,¹¹ Koran,¹² Deyo et al,¹³ Pecoraro et al,¹⁴ and Deyo.¹⁵

*The kappa statistic quantifies agreement between two observers or replicates measurements after adjusting for chance agreements.

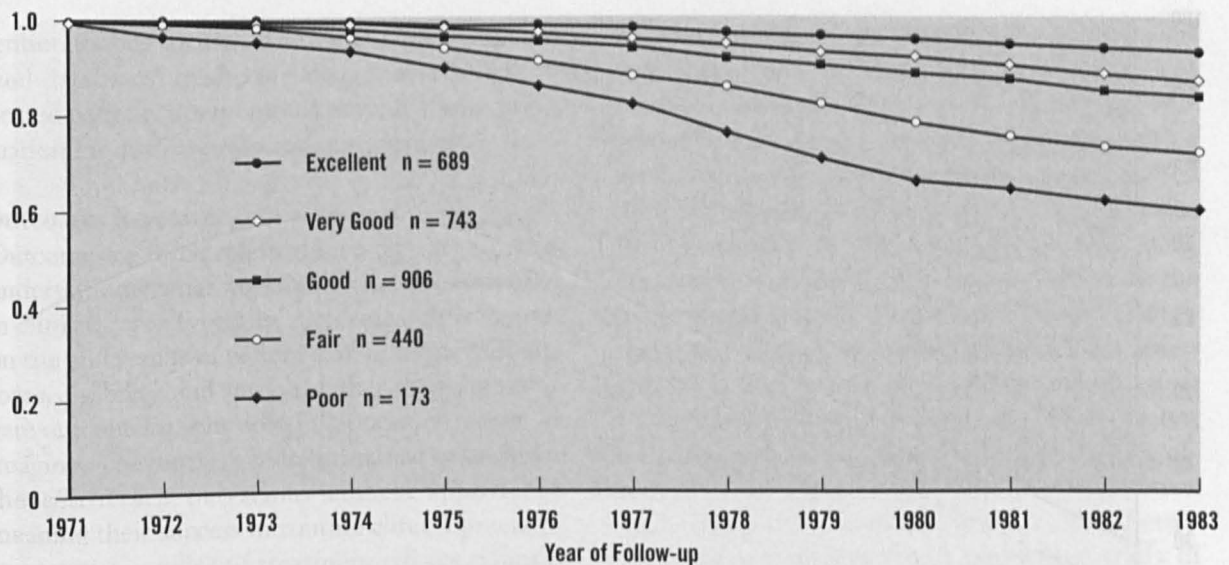


Figure 1. Male mortality in NHANES-I Epidemiologic Follow-up Study by levels of self-rated health.

Data from Idler and Angel.¹⁶

example, outcome measures might be used for quality-improvement purposes and to evaluate changes in the organization or content of clinical care. This use of outcome measurement is illustrated by the case example that began this article. Use of outcome measures for such purposes is increasingly advocated by accrediting organizations that deal with large ambulatory care systems.

In some cases hospitals or health plans have advocated using outcome measures to compare individual physicians or different treatments. For example, a hospital in Portland, Ore, used a popular health status questionnaire, the SF-36, to examine outcomes of hip-replacement surgery.¹⁷ As shown in Figure 2, these studies permitted comparisons between individual orthopedic surgeons and between different prosthetic devices. Several points were apparent from these studies. First, rapid improvement following surgery was generally observed, although the longer term results were somewhat less favorable. Second, although there were differences between surgeons and between prosthetic devices, at least some of the differences could be accounted for by differences in the baseline severity of the patients' conditions, as suggested by differences in SF-36 scores at the preoperative measurement. Nonetheless, it is easy to visualize how data of this sort might allow individual providers to compare notes or to choose a standard approach that appears to optimize patient outcomes.

Outcome data might also be used to evaluate

large system changes by a hospital, a health care system, or new health policy. If, for example, a health care system chose to alter its mix of generalist and specialist physicians, it might ask whether the end results in terms of patient outcomes were better or worse following the change. At one time, Washington State anticipated a major health policy reform (including universal coverage in a system similar to the Clinton plan), which ultimately was reversed by the state legislature. Had the plan gone forward, the state Department of Health anticipated developing a statewide outcome tracking system to help determine whether the overall health impact on the citizens of Washington State was favorable, unfavorable, or neutral.

Better data systems are necessary to apply outcome data to routine care. Some have suggested that brief health status questionnaires could be used routinely at clinic visits, much as vital signs are routinely measured.¹⁸ Nonetheless, manipulating such data, having them entered into a computerized database, and analyzing the results, all impose important burdens and costs. Although many health care providers have information systems that routinely collect information on patient use of services, amounts billed, diagnoses, and insurance coverage, these information systems typically have no data concerning patient symptoms, dysfunction, or satisfaction; nor do they have the kind of clinical detail that would be necessary to interpret these outcomes. Thus, realizing the full potential of outcome measures will likely require greater at-

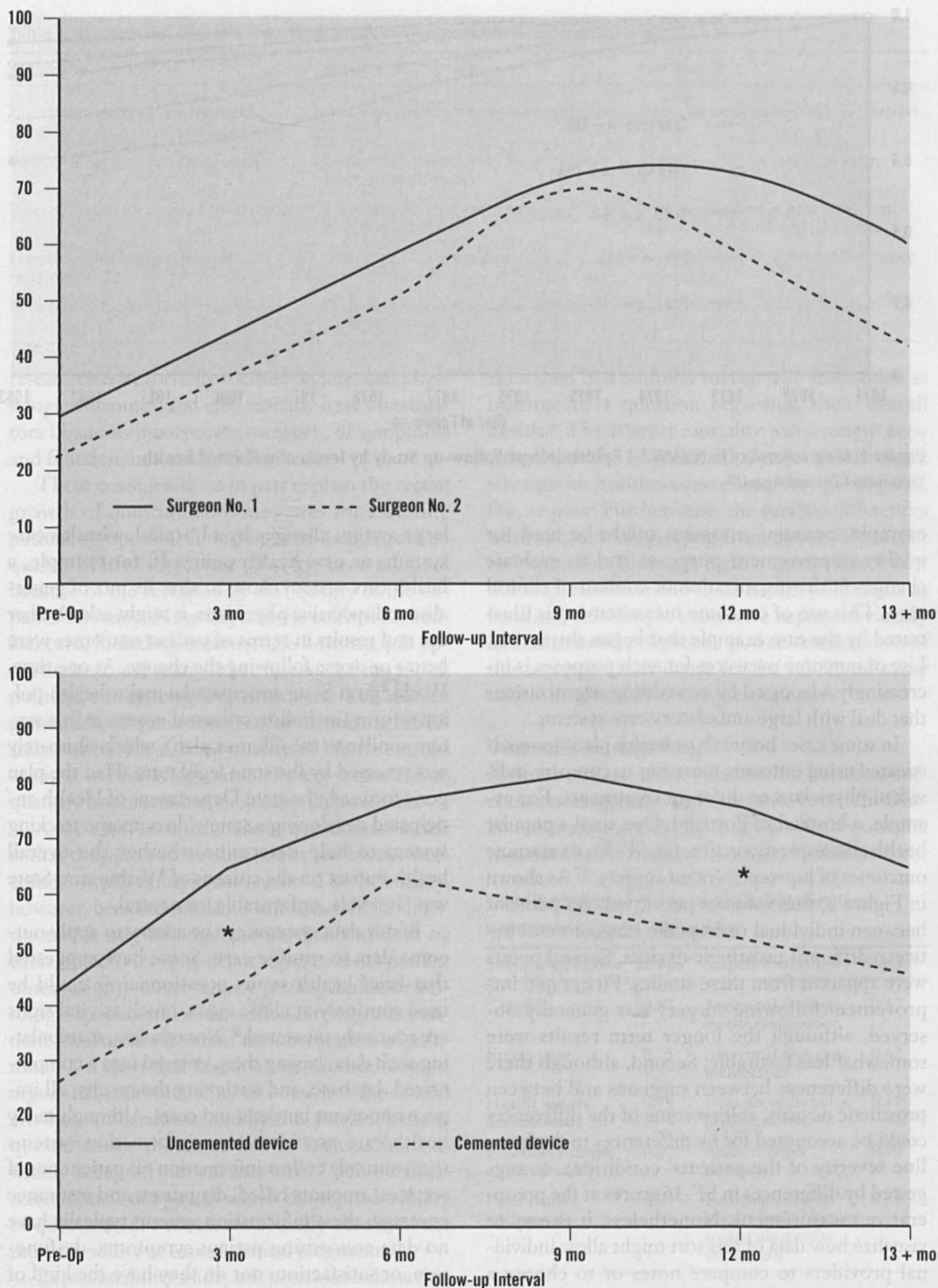


Figure 2. Average physical function scores of patients having hip-replacement surgery. Top: patient scores comparing surgeons. Bottom: patient scores comparing devices.

Data from Lansky et al.¹⁷ *Indicates differences significant at $P = 0.05$ level.

tention to their incorporation into large computerized databases. In the meantime, even small repeated patient surveys might provide useful information for quality-improvement purposes.¹⁹

Outcomes Research

Outcomes research refers to investigation aimed at understanding what works and what does not work in clinical care. Typically, such research is focused on the end results of patient care in terms of symptoms, disability, and survival rather than the surrogate outcomes of physiology, laboratory results, or imaging. The term is generally applied to studies of the *effectiveness* of various clinical approaches, meaning their success in routine clinical practices. In contrast, studies of treatment *efficacy* typically are randomized clinical trials conducted among highly selected patient populations, often in academic centers, often with leading experts providing the care and with closer monitoring and follow up than would generally be the case in routine settings. Efficacy studies address whether a treatment can work under ideal circumstances, not whether it does work as generally applied in routine care. Effectiveness in routine care is a function of efficacy, but also of diagnostic accuracy, physician skill in applying a treatment, patient compliance, and perhaps other factors that are artificially optimized in the clinical trial setting. Some would include randomized trials of treatment impact under the rubric of outcomes research if the trials incorporated patient-relevant outcome measures and were conducted in routine clinical settings.

More typically, however, outcomes research refers to the analysis of large administrative databases, such as insurance claims, to learn something about the costs, complications, and use of services associated with certain clinical strategies, or to observational studies, such as cohort studies, of patients receiving different treatments for the same condition but with the treatments determined by the course of usual care, rather than the intervention of a randomization schedule or investigator. Outcomes studies using these designs have in some cases added greatly to the understanding of the patient experience of various outcomes, unexpected consequences of therapy, and important gaps in clinical knowledge.

A series of studies conducted by the University of Washington Patient Outcome Research Team on low back pain can serve as examples. Back pain

was chosen for study because of its frequency, high costs, and wide treatment variability in styles of practice. One aspect of the project focused on outcomes of lumbar spine surgery, beginning with a synthesis of published literature on outcomes.²⁰⁻²² This synthesis showed that there is fair evidence from randomized trials to support the efficacy of conventional lumbar discectomy, but not the newer percutaneous techniques.²⁰ The cost-effectiveness of discectomy appeared to compare favorably with other widely accepted treatments, such as treatment of mild hypertension.²³ The evidence on spinal fusion suggested that this procedure offers little advantage when used in conjunction with simple discectomy for patients with herniated discs. There were no controlled trials of spinal fusion for degenerative discs in the absence of herniation.²²

Subsequent analyses of insurance claims databases showed that lumbar spine operations involving fusion procedures were associated with higher costs and complication rates than discectomy or laminectomy alone.^{24,25} These analyses also showed that reoperation rates following spinal fusion were at least as high as reoperation rates following laminectomy or discectomy without fusion. In the absence of randomized trials, these studies provided important descriptive information that supplemented the results of the literature syntheses.

The Patient Outcome Research Team also conducted a prospective cohort study that tracked outcomes of patients with sciatica or spinal stenosis cared for in the offices of orthopedic surgeons, neurosurgeons, or occupational medicine physicians. These data provided richer detail regarding patient outcomes than was previously available, and, as did the only randomized trial, suggested that after controlling for important confounding variables, surgery offered an advantage in many 1-year outcomes for patients with herniated discs and spinal stenosis.²⁶⁻²⁸ Return to work outcomes after 1 year and 3 years, however, were virtually the same with or without surgery.

The team then used the outcome data from the literature syntheses and claims analyses to develop an interactive computer-based video program for patients considering low back surgery. The intent of the program was to provide the best available information on outcomes of surgical or nonsurgical care tailored specifically to diagnosis and patient

age.²⁹ Thus, the program provides outcome data to patients in an effort to involve them more directly in decisions about their own care. Some data from these studies were also incorporated into clinical guidelines for managing low back pain.³⁰ These guidelines have had important impacts on quality and costs of care in some large health care organizations, and the research data might have contributed (with several other factors) to a national decrease in spinal fusion rates in 1994.

Commonly Used Instruments for Studying Health Outcomes

A wide variety of questionnaires is available for measuring symptoms and functional outcomes. These instruments typically measure outcomes in separate dimensions, such as physical functioning, emotional functioning, and role functioning. Some of these instruments are generic, meaning that they could be used for a wide variety of patients with a wide variety of conditions.³¹ Examples include the SF-36 (short form with 36 items, adapted from the Medical Outcomes Study)³² and its shorter version, the SF-12.³³ These instruments are currently in wide use. Other examples are the Sickness Impact Profile,³⁴ the Duke University Health Profile,^{35,36} the Dartmouth COOP charts,³⁷ and the Quality of Well-Being Scale.³⁸ The latter instrument and the EuroQoL (European Quality of Life Scale)³⁹ not only measure patient symptoms and functioning but also attach preference weights to each outcome state derived from interviews with large numbers of lay persons. These instruments permit the calculation of a single utility score for any outcome condition, which could theoretically be used in decision analysis and cost-effectiveness analysis.

In contrast to generic health-related quality-of-life instruments are disease-specific questionnaires, which focus more closely on the types of symptoms and dysfunctions that occur with particular disease conditions.³¹ Examples are available for studying the outcomes of asthma,⁴⁰ back pain,^{41,42} heart disease,^{43,44} arthritis,^{45,46} diabetes,⁴⁷ and many other conditions. Examples of studies using such instruments are cohort studies of back pain outcomes,^{27,28} studies of outcomes of different treatments for hypertension,⁴⁸ studies of outcome for rheumatoid arthritis,⁴⁹ treatment for acquired immunodeficiency syndrome,⁵⁰ and many others.

Generic questionnaires are needed when comparing the impacts of treatments for different diseases, including their cost-effectiveness. Disease-specific questionnaires are often more sensitive to subtle but clinically important changes and can pick out very specific health improvements that would not be apparent with a generic instrument. For research purposes, many investigators advocate including both a generic and a disease-specific measure.

Barriers and Hazards of Outcomes Analysis ***Unfair Outcome Comparisons***

A major problem with the use of outcomes for comparing providers or health care systems is that there might be incentives to "game" the results. One destructive way would be if certain providers simply declined to take the most difficult or severely ill patients. Accepting only healthier patients at the beginning assures having healthier patients at the end of treatment, and it has been suggested, for example, that cardiac surgeons might shun the tough cases.

More generally, making fair comparisons among physicians or hospitals based on their outcomes requires careful adjustment for case mix, so that providers with sicker or demographically disadvantaged patients are not unfairly penalized. Strategies for adjusting risk have been developed, generally incorporating at least patient age, sex, and comorbid conditions, all of which are available in most large automated databases.⁵¹ In some cases, adjustment for disease severity or complications is included, though such information is less consistently available. If health status or quality-of-life questionnaires were in widespread routine use, they could offer a powerful additional method for adjusting for baseline characteristics in addition to providing measures of outcome. Such adjustment techniques can help to level the playing field, but it is unlikely they will ever completely adjust for important differences among populations (which is, indeed, the reason for the ascendancy of random allocation in comparing treatment efficacy).

Factors Other Than Medical Management That Affect Outcomes

Many factors other than medical care affect the outcomes of a particular illness. For example, patients with multiple comorbid chronic diseases are likely to have worse outcomes than patients with

just a single condition. Homeless persons might have worse outcomes from many diseases than might patients who are more affluent. Patients who are compliant with drug therapy or with physical treatments are likely to have better success than those who are not. Genetic endowment can have important influences on outcomes regardless of the quality of care. Thus, it is important to define outcomes that are truly sensitive to the quality and effectiveness of the medical care delivered.

Difficulties in Measurement

For many conditions, the optimal timing or duration of follow-up for outcome assessments remains unclear. The timing can be quite different for acute and chronic diseases and even among chronic conditions. Different measures might be necessary for different settings and different populations, providing challenges for decision makers and for efforts to standardize measurement.⁵² Patient reading ability and language fluency have important effects on the feasibility of measuring health-related quality of life. Similarly, cognitive impairments can make it difficult or impossible to collect data from certain patients, and the validity of surrogate responses is variable.

Costs and Burdens

The costs associated with collecting and analyzing outcomes data can be substantial and might not be offset by immediate or obvious savings in other areas. The respondent burden for patients can also be substantial, depending on the length or frequency of measures.

Conclusions

The variations in care and cost pressures that have combined to create ever-increasing calls for accountability on the part of the health care professions has required not simply attention to the process of care, but to the end results. There has been a growing recognition that physiologic, laboratory, and imaging outcomes are sometimes poorly associated with symptomatic, functional, and survival outcomes, and thus the latter must be measured directly.

Symptoms, function, and quality of life can be quantified in a meaningful way, and a great variety of well-validated instruments are available for use. Their adoption and widespread use in routine care

settings for outcomes management will require far better data systems than are currently available and will require additional resources. In the research arena, one can hope that patient-centered outcomes will be increasingly incorporated into clinical trials. Outcomes research in the traditional sense, including large database analysis and cohort studies, complement, rather than compete with, randomized controlled trials. Finally, clinicians and investigators should be aware that outcomes research can in some cases precipitate political, legal, and media attacks and controversies, because such research focuses on clinical strategies that already have some credibility and a share in the marketplace.⁵³

Outcomes assessment cannot provide all the answers because of the many problems and pitfalls described above. Nonetheless, the current focus on outcomes can substantially improve both the quality of care and the quality of research in the US health care system.

References

1. Geyman JP. Evidence-based medicine in primary care: an overview. *J Am Board Fam Pract* 1998;11:46-56.
2. Wennberg JE, McPherson K, Caper P. Will payment based on diagnosis-related groups control hospital costs? *N Engl J Med* 1984;311:295-300.
3. Cherkin DC, Deyo RA, Wheeler K, Ciol MA. Physician variation in diagnostic testing for low back pain. Who you see is what you get. *Arthritis Rheum* 1994;37:15-22.
4. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med* 1991;324:781-8.
5. Oliver MF, Heady JA, Morris JN, Cooper J. WHO cooperative trial on primary prevention of ischemic heart disease using clofibrate to lower serum cholesterol: mortality follow-up. *Lancet* 1980;2:379-85.
6. Dwosh IL, Giles AR, Ford PM, Pater JL, Anastassiades TP. Plasmapheresis therapy in rheumatoid arthritis. A controlled, double-blind, crossover trial. *N Engl J Med* 1983;308:1124-9.
7. Lepor H, Williford WO, Barry MJ, Brawer MK, Dixon CM, Gormley G, et al. The efficacy of terazosin, finasteride, or both in benign prostatic hyperplasia. Veterans Affairs Cooperative Studies Benign Prostatic Hyperplasia Study Group. *N Engl J Med* 1996;335:533-9.
8. Nouwen A. EMG biofeedback used to reduce standing levels of paraspinal muscle tension in chronic

- low back pain. *Pain* 1983;17:353-60.
9. Feinstein AR. Clinical biostatistics. XLI. Hard science, soft data, and challenges of choosing clinical variables in research. *Clin Pharmacol Ther* 1977;22:485-98.
 10. Deyo RA, Andersson G, Bombardier C, Cherkin DC, Keller RB, Lee CK, et al. Outcome measures for studying patients with low back pain. *Spine* 1994;19(18 Suppl):2032S-36S.
 11. Wood RW, Diehr P, Wolcott BW, Slay L, Tompkins RK. Reproducibility of clinical data and decisions in management of upper respiratory illnesses: a comparison of physician and non-physician providers. *Med Care* 1979;17:767-79.
 12. Koran LM. The reliability of clinical methods, data, and judgments. *N Engl J Med* 1975;293:642-6, 695-701.
 13. Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain? *JAMA* 1992;268:760-5.
 14. Pecoraro RE, Inui TS, Chen MS, Plorde DK, Heller JL. Validity and reliability of a self-administered health history questionnaire. *Public Health Rep* 1979;94:231-8.
 15. Deyo RA. Measuring the functional status of patients with low back pain. *Arch Phys Med Rehabil* 1988;69:1044-53.
 16. Idler EL, Angel RJ. Self-rated health and mortality in the NHANES-I Epidemiologic Follow-up Study. *Am J Public Health* 1990;80:446-52.
 17. Lansky D, Butler JB, Waller FT. Using health status measures in the hospital setting: from acute care to "outcomes management." *Med Care* 1992;30(5 Suppl):MS57-73.
 18. Ellwood PM. Shattuck lecture—outcomes management. A technology of patient experience. *N Engl J Med* 1988;318:1549-56.
 19. Langley GJ, Nolan KM, Nolan TW, Norman CL, Provost LP. The improvement guide: a practical approach to enhancing organizational performance. San Francisco: Jossey-Bass, 1995:109-12.
 20. Hoffman RM, Wheeler KJ, Deyo RA. Surgery for herniated lumbar discs: a literature synthesis. *J Gen Intern Med* 1993;8:487-96.
 21. Turner JA, Ersek M, Herron L, Deyo R. Surgery for lumbar spinal stenosis. Attempted meta-analysis of the literature. *Spine* 1992;17:1-8.
 22. Turner JA, Ersek M, Herron L, Haselkorn J, Kent D, Ciol MA, et al. Patient outcomes after lumbar spinal fusions. *JAMA* 1992;268:907-11.
 23. Malter AD, Larson EB, Urban N, Deyo RA. Cost-effectiveness of lumbar discectomy for the treatment of herniated intervertebral disc. *Spine* 1996;21:1048-54.
 24. Deyo RA, Cherkin DC, Loeser JD, Bigos SJ, Ciol MA. Morbidity and mortality in association with operations on the lumbar spine. The influence of age, diagnosis, and procedure. *J Bone Joint Surg Am* 1992;74:536-43.
 25. Deyo RA, Ciol MA, Cherkin DC, Loeser JD, Bigos SJ. Lumbar spinal fusion. A cohort study of complications, reoperations, and resource use in the Medicare population. *Spine* 1993;18:1463-70.
 26. Keller RB, Atlas SJ, Singer DE, Chapin AM, Mooney NA, Patrick DL, et al. The Maine Lumbar Spine Study, Part I. Background and concepts. *Spine* 1996;21:1769-76.
 27. Atlas SJ, Deyo RA, Keller RB, Chapin AM, Patrick DL, Long JM, et al. The Maine Lumbar Spine Study, Part II. 1-year outcomes of surgical and non-surgical management of sciatica. *Spine* 1996;21:1777-86.
 28. Atlas SJ, Deyo RA, Keller RB, Chapin AM, Patrick DL, Long JM, et al. The Maine Lumbar Spine Study, Part III. 1-year outcomes of surgical and non-surgical management of lumbar spinal stenosis. *Spine* 1996;21:1787-95.
 29. Spunt BS, Deyo RA, Taylor VM, Leek KM, Goldberg HI, Mulley AG. An interactive videodisc program for low back pain patients. *Health Educ Res* 1996;11:535-41.
 30. Bigos S, Bowyer O, Braen G, Brown K, Deyo R, Haldeman S, et al. Acute low back problems in adults. Clinical practice guideline no. 14. Rockville, Md: Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services, December 17, 1994. (AHCPR publication no. 95-0642.)
 31. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;27:S217-32.
 32. Ware JE, Sherbourne CD. The MOS 36-item short-form survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
 33. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220-33.
 34. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981;19:787-805.
 35. Parkerson GR Jr, Broadhead WE, Tse CK. The Duke Health Profile. A 17-item measure of health and dysfunction. *Med Care* 1990;28:1056-72.
 36. Blake RL Jr, Vandiver TA. The reliability and validity of a ten-item measure of functional status. *J Fam Pract* 1986;23:455-9.
 37. Nelson EC, Landgraf JM, Hays RD, Wasson JH, Kirk JW. The functional status of patients. How can it be measured in physicians' offices? *Med Care* 1990;28:1111-26.
 38. Kaplan RM, Bush JW, Berry CC. Health status; types of validity and the Index of Well-Being.

- Health Serv Res 1976;11:478.
39. EuroQoL Group. EuroQoL - a new facility for the measurement of health-related quality-of-life. *Health Policy* 1990;16:199-208.
 40. Juniper EF, Guyatt GH, Ferrie PJ, Griffith LE. Measuring quality of life in asthma. *Am Rev Respir Dis* 1993;147:832-8.
 41. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983;8:141-4.
 42. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. *Spine* 1995;20:1899-909.
 43. Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Prodzinski J, McDonell M, et al. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. *J Am Coll Cardiol* 1995;25:333-41.
 44. Guyatt GH, Nogradi S, Halcrow S, Singer J, Sullivan MJ, Fallen EL. Development and testing of a new measure of health status for clinical trials in heart failure. *J Gen Intern Med* 1989;4:101-7.
 45. Fries JF, Spitz PW, Kraines RG, Holman HR. Measure of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
 46. Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis. The arthritis impact measurement scales. *Arthritis Rheum* 1980;23:146-52.
 47. Boyer JG, Earp AL. The development of an instrument for assessing the quality of life of people with diabetes. *Diabetes* 1997;35:440-53.
 48. Beto JA, Bansal VK. Quality of life in treatment of hypertension. A metaanalysis of clinical trials. *Am J Hypertens* 1992;5:125-33.
 49. Bombardier C, Ware J, Russell IJ, Larson M, Chalmers A, Read JL. Auranofin therapy and quality of life in patients with rheumatoid arthritis. Results of a multicenter trial. *Am J Med* 1986;81:565-78.
 50. Weissman JS, Cleary PD, Seage GR 3rd, Gatsonis C, Haas JS, Chasan-Taber S, et al. The influence of health-related quality of life and social characteristics on hospital use by patients with AIDS in the Boston Health Study. *Med Care* 1996;34:1037-54.
 51. Wray NP, Hollingsworth JC, Peterson NJ, Ashton CM. Case-mix adjustment using administrative databases: a paradigm to guide future research. *Med Care Res Rev* 1997;54:326-56.
 52. Deyo RA, Patrick DL. Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Med Care* 1989;27(3 Suppl): S254-68.
 53. Deyo RA, Psaty BM, Simon G, Wagner EH, Omenn GS. The messenger under attack: intimidation of researchers by special-interest groups. *N Engl J Med* 1997;336:1176-80.