POLICY FORUM: HEALTH CARE POLICY

Outcomes Research: Measuring the End Results of Health Care

Carolyn M. Clancy and John M. Eisenberg

Substantial variations in medical practice have been well documented in the United States and abroad. Patients with the same condition are treated quite differently in different locations, irrespective of organizational and financial arrangements. This lack of uniformity in medical practice has stimulated extensive inquiry into the relation between the use of clinical services and their end results (*1*, *2*).

Associating differences in the process of care with differences in outcomes can clarify which services are worth providing, which services represent misused resources or a need for more evidence about their effectiveness, and where clinicians and organizations have opportunities for improvement (*2*). Policy-makers and purchasers are interested in identifying sources of cost without benefit. Clinicians need to select effective treatments, and patients want to make informed treatment choices (*3*). All these concerns have stimulated researchers in the clinical, quantitative, behavioral, and social sciences to expand the methods and metrics used to evaluate the effects of health services. Outcomes research—the study of the end results of health services that takes patients' experiences, preferences, and values into account—is intended to provide scientific evidence relating to decisions made by all who participate in health care.

Measuring Outcomes

Clinical success has traditionally been appraised in terms of mortality, physiological measures such as blood pressure or diagnostic test results that are surrogates for physiologic function (such as laboratory tests, radiographic findings, or biopsy results), and definable clinical events. Clinical trials have produced these objective measures as their primary dependent variables. Seldom have patients' preferences for outcomes and risks of treatment been used to evaluate health services; they often have been perceived as important but subjective and unreliable. However, patients and clinicians must increasingly make decisions associated with different types of outcomes, such as length of survival, preservation of function, or pain relief (*4*, *5*).

The dimensions of health and well-being that encompass consequences for the daily lives of individual patients are referred to as health-related quality of life (HRQL). Broad aspects of HRQL include health perceptions, symptoms, functioning, and patients' preferences and values. The sum of these constitutes a continuum of effects of health care services on health and well-being, ranging from mortality to patient satisfaction.

Health perceptions. An individual's rating of overall health is among the best predictors of mortality and future use of services (*6*). One of the best-established measures of health perceptions is the patient's symptoms. Validated symptom inventories with standardized questions, some of which clinicians have asked of patients for centuries, allow comparisons among individuals or groups. For example, the American Urological Association (AUA) Symptom Index scale assesses the frequency and severity of symptoms produced by benign prostatic hyperplasia (BPH) and is now used by over 80% of practicing urologists (*7*).

Functional measures. These can be used to assess the net impact of health services on overall or general health, as well as the impact on a particular disease. These measures assess the ability of individuals to carry out daily activities that are important to them, ranging from general activities of daily life to functions specific to a particular organ or body part. Although existing tools vary in comprehensiveness, an emerging consensus among developers and users is that they should include physical function, mental function or psychological distress, limitations in social and role function due to health problems, and general health perceptions (*8*). Items on these scales are usually objective questions, such as "Can you walk a block?" or "How often in the past month have you felt sad or blue?" The SF-36, a brief questionnaire derived from a much longer instrument developed to assess the effects of varying levels of health insurance, is widely used in clinical trials and in clinical practice.

Preference-based measures. Standardized inquiries about functional status complement the information obtained from the more traditional measurements of health, but they offer few insights about the meaning of health states to an individual's daily life. For example, the ability to walk a block will have different implications for an injured athlete than for a sedentary person. Preference-based outcome measures use a variety of techniques that ask individuals to make judgments about the value of health states. The Quality of Well-Being Scale, which includes explicit questions about social role function, typifies more global approaches to determining general health status (*9*). Alternatively, the disease-specific BPH Impact Index addresses the impact of prostatic symptoms; answers to these questions provide evaluative rather than purely descriptive information (*7*). Preference-based measures are particularly relevant to alternative interventions associated with little clinical uncertainty but possible striking differences in quality of life. For example, mastectomy and breast-sparing surgery for women with breast cancer are equally effective in reducing mortality, but women have different values for the outcomes of alternative treatments.

Patient satisfaction. This measure reflects more than technical aspects of care. Interpersonal aspects have been shown to influence adherence to recommended treatments and advice, as well as individuals' ability to manage their own conditions (*10*).

Several measures that incorporate both health outcomes and time in order to capture quantity and quality of life in a single metric have been developed and tested for different purposes. These include quality-adjusted life years (QALY), potential years of life lost, disability-adjusted life years (DALY) (*11*), health-adjusted life expectancy, and years of healthy life (*12*). These measures can be used to inform decision-makers about how to allocate health care resources to improve population health. International public health organizations are now considering the use of DALYs to evaluate the use of scarce resources for providing health care in developing nations. Use of these measures assumes both that quality of life can be measured accurately and that community preferences can represent individual preferences. Research that estimates community preferences and compares preferences of different communities is a high priority for the field.

C. M. Clancy is director of the Center for Outcomes and Effectiveness Research and J. M. Eisenberg is administrator of the Agency for Health Care Policy and Research, U.S. Department of Health and Human Services (HHS), 2101 East Jefferson Street, Rockville, MD 20852, USA. The views expressed here are those of the authors and do not necessarily represent the position of the Agency for Health Care Policy and Research or of HHS.

Which Outcome Measure to Use?

There is a wide array of tools that can be used to assess one or more dimensions of the health outcomes continuum. A significant advance within the past decade has been the development of psychometrically valid instruments that are short and easy to administer (*13*).

Although there is no precise estimate of the proportion of clinical trials that includes these measures, the peer-reviewed literature suggests that these measures are being included in clinical trials with increasing frequency (*5*). Apparent consensus regarding which measures to use has evolved slowly within selected clinical domains. Measures are usually selected on the basis of criteria that include prior use in a comparable patient population, respondent burden, specific hypotheses being tested, and investigators' experience.

Purchasers' demands for common measures of health care quality for the populations they serve have stimulated the development of measurement tools that can be broadly applied and linked with consumers' and purchasers' decisions. Current initiatives include the use of outcomes measures by purchasers and consumers to compare health plans, by states to assess hospitals' experience in performing cardiac procedures (*14*), by health plans to assess clinicians' performance (*15*), and by policy-makers to evaluate the performance of health plans in providing care to Medicare beneficiaries.

These extensions of research methods to the "real world" highlight some challenges for the field. An ongoing debate concerns when to use a general measure of health outcome and when a disease- or condition-specific measure is more appropriate. General measures can describe outcomes for people with a range of conditions and are especially useful for population surveys. However, the relevant outcome may be obscured if a general outcome measure is the only one used. Disease-, condition-, or population-specific measures are complementary to general measures because they usually assess outcomes that are temporally related to clinical decision-making, are more responsive to changes in health resulting from a particular disease, and may be perceived by clinicians to be more relevant (*16*). Specific instruments have been widely used to study arthritis and musculoskeletal conditions (*17*), chronic lung disease (*18*), and visual disorders (*19*), as well as to assess pain (*20*). Many clinical trials use both types of measures when feasible.

Additional issues include the impact of age and other social and demographic factors on responses, adjustment for severity of disease, and the relation between organizational characteristics of health systems and clinical care (*21*). Comparisons of health plans or hospitals that do not include corrections for severity of illness may inadvertently encourage plans and providers to avoid the sickest patients. Adjusting for differences in severity of a single condition is often done with predictive models derived from physiological measures (*22*). Innovative methods to clarify the magnitude, extent, and distribution of other risk factors (such as race, age, sex, culture, and socioeconomic status) associated with health outcomes are needed, because many measures have not been tested in diverse populations.

Future Needs and Directions

With a few notable exceptions, a growing appreciation of the need to incorporate patients' preferences and values in clinical decision-making has not been matched by widespread use of outcome measures in daily practice. There have been dramatic increases in the application of standardized assessments in response to the concerted efforts of a cadre of large employers, but current accreditation requirements are not required of all plans or in all markets (*23*). A challenge to the scope of current quality measures has been the compromises needed to balance what we want to measure with existing information systems. The development and use of electronic medical records in clinical practice will significantly enhance the quantity and quality of outcomes measures that can be used to assess clinical performance, for external comparison as well as for internal improvement.

Additional work to enhance the interpretability of outcome measures, particularly in terms of clinical significance, is needed to increase the usefulness of these tools (*24*). Clinicians are unlikely to use patient-reported outcome measures routinely unless the reports are as familiar to them as blood pressure and other physiologic measures. This cannot occur until outcomes measures are developed that are easy to include in daily practice. Increased availability of computerized information systems in typical practice settings will be essential. Crucial features of studies that have had an unambiguous impact on practice include the early involvement of the relevant medical professional organizations and opinion leaders in both research and dissemination.

Individual patients' interest in selecting treatments that are consonant with their preferences, and their ability to communicate those values, will also enhance wider use of patient-reported measures. Currently, patients are surveyed with increasing frequency about their experiences and satisfaction with care, and a growing number of patients now prepare for clinical encounters by first searching the Internet (*25*). Many of these efforts occur somewhat peripheral to clinician-patient encounters, and there is no easy way for patients to make choices of providers or treatments systematically. However, patients want this information, and a majority want to know how "patients like me" have fared in similar circumstances (*26*). Sophisticated information systems that can provide reports to clinicians and patients on physiological and patient-reported outcomes will enhance the inclusion and use of patient-reported measures and profoundly shape the opportunities for patients to be co-managers of their health and health care.

References and Notes

1. J. Wennberg and A. Gittelsohn, *Science* **182,** 1102 (1973); K. McPherson, J. E. Wennberg, O. B. Hovind, P. Clifford, *N. Engl. J. Med.* **307,** 1310 (1982); M. R. Chassin *et al.*, *JAMA* **258**, 2533 (1987); L. L. Leape *et al.*, *ibid.* **263**, 669 (1990); Dartmouth Medical School, *The Dartmouth Atlas of Health Care* (American Hospital Publishing, Chicago, 1996).
2. "Health Care's Weird Geography," *New York Times*, 25 October1997, p. A10.
3. C. Laine and F. Davidoff, *JAMA* **275**, 152 (1996).
4. B. J. McNeil, R. Weichselbaum, S. G. Pauker, *N. Engl. J. Med.* **305**, 982 (1981); J. E. Wennberg, *Inquiry* **31**, 296 (1994).
5. G. H. Guyatt, D. H. Feeny, D. L. Patrick, *Ann. Intern. Med.* **118**, 622 (1993).
6. G. A. Kaplan and T. Camacho, *Am. J. Epidemiol.* **117**, 292 (1983); J. E. Connelly, J. T. Philbrick, G. R. Smith, D. L. Kaiser, A. Wymer, *Med. Care* **27**, S99 (1989).
7. M. J. Barry *et al.*, *Med. Care* **33**, AS145 (1995).
8. J. E. Ware Jr., *Annu. Rev. Public Health* **16**, 327 (1995).
9. M. M. Chenn, J. W. Bush, D. L. Patrick, *Policy Sci.* **6**, 71 (1975); R. M. Kaplan and J. P. Anderson, *Health Serv. Res.* **23** 203 (1988).
10. S. Greenfield, S. Kaplan, J. E. Ware Jr., *Ann. Intern. Med.* **102**, 520 (1985); S. Greenfield, S. H. Kaplan, J. E. Ware Jr., E. M. Yano, H. J. Frank, *J. Gen. Intern. Med.* **3**, 448 (1988).
11. C. Murray and A. Lopez, *Science* **274**, 740 (1996).
12. M. R. Gold *et al.*, in *Cost-Effectiveness in Health and Medicine*, M. R. Gold, J. E. Siegel, L. B. Russell, M. C. Weinstein, Eds. (Oxford Univ. Press, New York, 1996), pp. 82–123.
13. C. A. McHorney, *Ann. Intern. Med* .**127**, 743 (1997).
14. M. R. Chassin, E. L. Hannan, B. A. DeBuono, *N. Engl. J. Med.* **334**, 394 (1996).
15. A. M. Epstein, *JAMA* **279**, 1691 (1998).
16. M. H. Brenner, B. Curbow, M. W. Legro, *Med. Care* **33 (Suppl.)**, AS236 (1995).
17. R. F. Meenan, P. M. Gertman, J. H. Mason, *Arthritis Rheum.* **23**, 146 (1980).
18. G. H. Guyatt *et al.*, *Can. Med. Assoc. J.* **15**, 919 (1985).
19. E. Steinberg *et al.*, *Arch. Ophthalmol.* **112**, 630 (1994).
20. I. McDowell and C. Newell, *Measuring Health* (Oxford Univ. Press, New York, ed. 2, 1996).
21. B. E. Landon, I. B. Wilson, P. D. Cleary, *JAMA* **279**, 1377 (1998).
22. L. I. Iezzoni, *J. Invest. Med.* **43**, 136 (1995).
23. J. K. Iglehart, *N. Engl. J. Med.* **335**, 995 (1996).
24. R. Deyo and D. Patrick, *Med. Care* **33**, AS286 (1995).
25. D. Blumenthal, *JAMA* **278**, 1622 (1997).
26. S. Robinson and M Brodie, *Jt. Comm. J. Qual. Improv.* **23**, 239 (1997).