Comparing healthcare outcomes

Carol Orchard

Governments are increasingly concerned to compare the quality and effectiveness of healthcare interventions but find this a complex matter. Crude hospital statistics can be dangerously misleading and need adjusting for case mix, but identifying and weighting the patient characteristics which affect prognosis are problematical for conceptual, methodological, and practical reasons. These include the inherently uncertain nature of prognosis itself and the practical difficulties of collecting and quantifying data on the outcomes of interest for specific healthcare interventions and known risk factors such as severity.

Need for outcome measures

The combined pressures of "medical inflation," fiscal constraints, and a shift in attitudes towards publicly funded services during the 1980s have made the search for measures of quality, efficiency, and effectiveness in health care a government priority in industrialised countries. Providers of health services are increasingly required to account for the resources they use. In order to do so they must devise ways of measuring the outcomes of their activities and the extent to which these meet specified objectives, be they medical, social, or financial.

The complexity of what hospitals do makes such measurement difficult even in the narrow specialty of acute care, not least because the resources used, the choice of treatments, and the observed results (Donabedian's structure, process, and outcome²) are so dependent on the sorts and conditions of patients admitted—the hospital "case mix." Attempts to measure this to date have focused largely on resource consumption, by collating diagnosis and procedure codes in order to form groups which are homogeneous in costs (proxied by length of hospital stay). However, if comparisons of performance between hospitals are to be meaningful, diagnosis related groups or their English equivalents, healthcare resource groups, will have to be supplemented by measures of input and output. In other words, data about patients will need to be made manageable by being grouped in terms of needs for treatment and the prognoses arising from it.

Casemix developers in the United States and elsewhere have found that data about patients (as opposed to treatments) are difficult to organise in this way, but one approach is to work from the assumption that patients' needs for health care are best defined in terms of the expected outcomes of specified treatment options. Developing casemix groups on these lines will then entail addressing the conceptual and practical problems surrounding (a) the nature of prognosis, (b) the sources of data about patients, (c) the statistical analysis of those data, (d) the identification and weighting of patient risk factors, and (e) the definition and measurement of severity.

Prognosis

The nature of prognosis itself is a major difficulty—firstly, because differential rates of illness and recovery are substantially associated with non-medical factors such as socioeconomic class, diet, smoking, family

Summary points

- Hospital activities are complex and difficult to measure and their outcomes data need adjusting for case mix
- Existing casemix systems focus on treatments, but outcomes data will need to focus on patients
- Prognosis itself is inherently uncertain and substantially affected by non-medical factors
- Medical prognoses are multidimensional, qualitative, time specific, and disease specific and may be difficult to attribute to specific treatments
- Using observational data for outcomes measurement means allowing for inaccuracy, bias, confounding, and chance
- Identifying and weighting patient characteristics associated with poor outcomes entail rigorous research
- Prognostic factors, including severity, are difficult to quantify
- The benefits of casemix adjusted outcomes data should outweigh the problems and costs in producing them

support, genetic predispositions, and physiological reserve. In other words, there is a need to distinguish between health outcomes and healthcare outcomes.³ Even for factors whose effects on general health outcomes are established little is known about their relative weight and the reasons they affect some people more than others. Secondly, prognosis for specific conditions, taking account of medical interventions, is unlikely ever to be other than uncertain and inherently probabilistic⁴ even if accurate data on all known risk factors could be collected. Finally, estimating these probabilities is complicated by the nature of the outcomes (see box).

Assessing medical outcomes

- Outcomes are multidimensional
- Most outcomes are qualitative
- Assessment of outcomes will be affected by timing
- Subgroups of diseases may have differing outcomes
- Outcomes may not be attributable to specific treatments

For example, prognosis is multidimensional. At least five aspects have been identified of even a restricted definition of "outcomes" which focuses on the negative results for patients—namely, death, disease, disability, discomfort, and dissatisfaction.⁵ Many outcomes are qualitative, raising the question of how and by whom they are to be quantified. Their assessment will be affected by the time period over which they are made. For instance, over what period of weeks or years should outcomes for hip replacements be evaluated? Outcomes are likely to vary across

National Casemix Office,
Information Management
Group of the NHS
Executive, Winchester,
Hampshire SO23 9JA
Carol Orchard, research
assistant

BMJ 1994;308:1493-6

subgroups of a disease and may or may not be attributable to specific treatments.⁴

Even if consensus can be reached about the anticipated results that particular interventions are intended to achieve variations between observed and expected outcomes are difficult to assess. Quality of care is commonly assessed on the basis of negative outcomes, but as death is unlikely in most cases, other, more ambiguous outcomes will have to be measured. Quantifying such qualitative concerns as pain or discomfort, loss of mobility, or other kinds of disablement is fraught with difficulty, and the methods by which this is done need to be rigorous—not least in making explicit the assumptions and values on which they are based.

Sources of data

More difficulties surround the sources from which patient related data are obtained. The only unambiguous method of equalising risk among groups of patients having different treatments is through randomised controlled trials.¹ The collection of data from these under the auspices of the Department of Health's research and development strategy⁶ and their publication as meta-analyses are most welcome. However, extensive use of randomised controlled trials is prohibited by practical difficulties (for instance, achieving true randomisation and the cost in time and money), doubts about the extent to which generalised conclusions about "typical" patients can be applied to particular groups (such as the very old or very young), and ethical problems surrounding informed consent and the withholding of beneficial or application of potentially harmful treatments.¹ Moreover, randomised controlled trials are likely to be better tests of effectiveness (whether the chosen treatment "works") than of quality (whether the treatment is being delivered properly).

In the United States the need for valid outcome measures is regarded as particularly urgent for economic reasons. The Office of Health Economics (personal communication) estimates that by 1992 healthcare expenditure accounted for over 14% of gross domestic product (more than twice the percentage of the United Kingdom), 40% of this being reimbursed by the government. Because of the limitations of randomised controlled trial data, resort has been made to routine observational data from hospital discharge summaries and health insurance claims for comparing performance between providers. These are

Sources of data on patient outcomes

- Results of randomised controlled trials ("gold standard")
- Meta-analysis (combined results of randomised controlled trials)
- Routine observational data (surveys, hospital discharge summaries, etc)
- Cross design synthesis (combined randomised controlled trial results and observational data)

adjusted for risk under the Health Care Financing Administration's prospective payment system based on diagnosis related groups (in some states refined by severity scoring, as diagnosis related groups explain resource use rather than prognosis).

However, as doubts are widespread about the accuracy of such data and the bias and confounding inherent in their use, the United States General Accounting Office has advocated an alternative source—namely, the so called cross design synthesis, which combines the results of randomised controlled trials with the contents of large databases. The intention is to compensate for the weaknesses of both sources of data by combining their strengths.⁷ Unfortunately, it is doubtful whether the problems associated with using observational data from routine sources can be overcome so easily.^{3,8}

More accurate assessments of prognosis and the risk factors concerned will call for additional data to be collected, either from medical records or from patient surveys. With this in mind many patient surveys have been designed, both generic (such as those using the SF 36 questionnaire and the Nottingham health profile) and disease specific (such as the quality of life index for cancer patients or the Stanford health assessment questionnaire for those suffering from arthritis), though more research is needed to validate them.⁹ Meanwhile, in the United Kingdom better access to clinical data is being brought nearer by the clinical terms project,¹⁰ Read codes, and other information management improvements taking shape under the information management and technology strategy of the NHS.¹¹

Statistical problems

Nevertheless, major practical difficulties need to be overcome before analyses based on observational data—whether or not combined with randomised controlled trials—could be regarded as credible. The first and most obvious problem, as noted above, is the accuracy of the data themselves. The coding, recording, and measurement of routine patient data in hospitals may be adequate for internal management but not for outcomes evaluation. Secondly, allowance must be made for random variations, which may account for much of the difference between providers' results. When numbers of cases are large enough it is possible to discover whether or not these differences are statistically significant, but when numbers are very small it is difficult to be sure that any variations are not entirely due to chance.¹²

Less tractably, observational data are inevitably prone to bias and confounding. Biases arise because in normal medical practice patients are not selected for treatment randomly (as they would be in randomised controlled trials) but on the basis of a range of factors to which different clinicians attach different weights. Decisions about referrals and treatments are made on the basis of individual prognostic assessments, which are likely to be influenced by non-medical factors as well as by estimates of disease severity, comorbidity,

Coding, recording, and measurement of routine patient data in hospitals may not be adequate for outcomes evaluation

ULRIKE PREUSS

and the anticipated effectiveness and acceptability of the intervention.

Confounding—that is, the presence of additional factors which render apparent associations spurious—is particularly difficult to eliminate, as several recent studies illustrate. For example, controversy surrounded the respective outcomes for transurethral resection of the prostate compared with open surgery for prostatectomy,^{13 14} and for angioplasty compared with coronary artery bypass grafting for patients with severe angina.^{15 16} In both cases it was argued that what appeared to be better outcomes for transurethral resection of the prostate and coronary artery bypass grafting had been confounded by patient factors—namely, age and frailty in one case and severity in the other. At the same time researchers need to beware of cancelling out the effects of one confounding variable by overadjusting for another. Data on respiratory illness might be adjusted for cigarette smoking, for instance, and mask the effects of socioeconomic class.

Risk adjustment

Notwithstanding these difficulties, the need for information about outcomes makes the trend towards using risk adjusted observational data for this purpose unstoppable. The research implications may be considerable, as the efforts devoted to researching the aetiology of diseases and the relative effectiveness of different treatments have not been matched by similarly extensive research into patient characteristics associated with good or bad outcomes.

To date, the growing body of published evidence on outcomes mainly reflects debates about conceptual and methodological matters. Most of the papers examining prognostic factors associated with specific diseases concern acute manifestations or surgical interventions. Comparatively few address the problem of chronic or intractable conditions such as AIDS, arthritis, or diabetes—perhaps understandably, as research into outcomes for chronic conditions like rheumatoid arthritis, for example, is bedevilled both by the problem of agreeing on and measuring the outcomes of interest¹⁷ and by the unpredictable nature of the disease itself, whose clinical course is unpredictable and improvements in which may occur regardless of therapy.¹⁸

For acute conditions there is evidence that the association between patient characteristics and results of treatments vary, not only between conditions but between outcomes. In the United States a study of Medicare patients undergoing elective cholecystectomy or transurethral resection of the prostate found that “the factors associated with death were not the same factors associated with adverse occurrence [complications] or failure to rescue [death after an adverse occurrence].”¹⁹

A similar conclusion about differential effects of patient risk factors was reached in another American study of five acute conditions—stroke, lung cancer, pneumonia, acute myocardial infarction, and congestive heart failure. It was found that blood urea nitrogen value, consciousness level, and systolic blood pressure were important key clinical findings for all five conditions, and arterial pH and respiratory rate were important for four of them. However about a fifth of the key clinical findings were condition specific. Researchers conclude that generic factors may be helpful in predicting hospital mortality but that they are not the sole predictors.^{12 20}

Published work suggests that the following factors are predictive for many acute conditions: age of the patient, severity of the illness at presentation, comorbidity, physiological test results, and previous history or frailty of the patient, or both. Their effects

are differential and may be difficult to quantify—even age, as there are suggestions that biological rather than chronological age is important.²¹ Severity (of secondary as well as primary diagnoses) is often significant, but its definition and measurement are problematical, as discussed below. Like age, comorbidity may or may not be relevant to the condition or treatment being assessed; moreover the recording of secondary diagnoses is notoriously difficult to standardise. Physiological tests will vary according to the condition (for example, antigens or nuclear shape of cancer cells for neoplasms, weight loss before surgery) and may need to be very detailed and specific. Finally, recording of previous history, like comorbidity, is difficult to standardise.

Severity

It seems incontrovertible that allowance must be made for severity when assessing healthcare outcomes, but doing this is far from straightforward. Firstly, clinicians’ subjective opinions of severity are not only important determinants of the type of care received but may also be prognostic factors in their own right because of their influence on interactions with patients. Secondly, failure to respond to treatment may be an indication of severity—or an indication that the chosen treatment is inappropriate. Thus for outcome measurements severity needs to be assessed solely in terms of patients—that is, excluding aspects of the structure or process of health care.³

Several severity scoring systems have been developed, the best known from the United States, but each uses a different definition of what has been described as an “important but nebulous concept.”²² These range from “the treatment difficulty presented to physicians due to the extent and interactions of a patient’s diseases”²³ to the risk of imminent death in an intensive care unit²⁴ or the risk of imminent organ failure.²⁵ Some systems correlate increasing case complexity with intensity and costs or, using cancer staging as a model, rate the relevant disease rather than the patient.²⁶

Some of these severity measures are computerised and use retrospective data from hospital discharge summaries. This has obvious drawbacks if the intention is to assess the quality of care, not least because it becomes difficult to separate iatrogenic events from those occurring because of the disease process. The prospective and exogenous (that is, patient rather than treatment focused) systems entail collecting additional data from medical records. This will mean extra expense and compound the difficulties of achieving quality and consistency of data within and between hospitals.

Conclusions

The problems discussed above are a formidable list. Nevertheless, in the present philosophical and economic setting they are likely to be viewed as challenges rather than insurmountable barriers. Clearly casemix comparisons based on similar needs for care and expected prognoses are essential if purchasers of health care are to assess the appropriateness of existing referral and treatment patterns and make decisions based on quality and equity as well as cost. However, the benefits could be much wider. Doctors and patients would surely gain from having manageable data about the outcomes associated with particular prognostic factors, so that discussions about treatment options could take place on the basis of shared knowledge about probable risks.

Realising these benefits will entail balancing the ease of collection and doubtful validity of generalised

Common patient risk factors in acute illnesses

- Age
- Severity
- Comorbidity
- Physiological findings
- Previous history or frailty

routine data against the better prognostic estimations and greater costs of disease specific physiological data.²⁷ Until more is known about costs, benefits, and valid methods it will be advisable to proceed with caution, building up information about prognostic groupings condition by condition and using empirical data from medical audit systems and outcomes research to validate them.

While this is taking place the conclusion reached in the United States, notwithstanding the sums already spent on outcomes research there, is that for the time being risk adjusted outcomes data are best used for quality management by hospitals themselves, flagging up areas for internal investigation. Publication of crude mortality figures inadequately adjusted for risk would be counter productive if it resulted here, as it did in the United States, in the most egregious rates being found at a hospice for terminally ill patients somewhere in the West.²⁸

I thank Mr Trevor Sheldon, of the Centre for Health Economics, University of York, and Drs Phil Anthony, Siva Prakash, and Hugh Sanderson, of the National Casemix Office, for helpful comments.

- 1 Cochrane AL. *Effectiveness and efficiency: random reflections on health services*. London: Nuffield Provincial Hospitals Trust, 1972.
- 2 Donabedian A. The quality of care: how can it be assessed? *JAMA* 1988;260:1743-8.
- 3 Gulliford MC. Evaluating prognostic factors: implications for measurement of health care outcome. *J Epidemiol Community Health* 1992;46:323-6.
- 4 Fries JF, Ehrlich GE. *Prognosis: contemporary outcomes of disease*. Bowie, MD: Charles Press, 1981.
- 5 Holland WW, ed. *Evaluation of health care*. Oxford: Oxford University Press, 1983.
- 6 Peckham M. *Research for health*. London: Department of Health, 1993.
- 7 United States General Accounting Office. *Cross design synthesis: a new strategy for medical effectiveness research*. Washington, DC: GAO, 1992.
- 8 Cross design synthesis: a new strategy for studying medical outcomes? *Lancet* 1992;340:944-5.

- 9 McDowell I, Newell C. *Measuring health: a guide to rating scales and questionnaires*. Oxford: Oxford University Press, 1987.
- 10 Buckland R. The language of health: a clinical language underlies the NHS information strategy. *BMJ* 1993;306:287-8.
- 11 Leaning MS. The new information management and technology strategy of the NHS: person centred. *BMJ* 1993;307:217.
- 12 Iezzoni LI, Ash AS, Coffman GA, Moskowitz MA. Predicting in-hospital mortality: a comparison of severity measurement approaches. *Med Care* 1992;30:347-59.
- 13 Fisher ES, Wennberg JE. Administrative data in effectiveness studies: the prostatectomy assessment. In: Heithoff KA, Lohr KN, eds. *Effectiveness and outcomes in health care*. Washington, DC: National Academy Press, 1990.
- 14 Concato J, Horwitz RJ, Feinstein AR, Elmore JG, Schiff SF. Problems of comorbidity in mortality after prostatectomy. *JAMA* 1992;267:1077-82.
- 15 Hartz AJ, Kuhn EM, Pryor DB, Krakauer H, Young M, Heudebert G, et al. Mortality after coronary angioplasty and coronary bypass surgery (the national Medicare experience). *Am J Cardiol* 1992;70:179-85.
- 16 RITA Trial Participants. Coronary angioplasty versus coronary artery bypass surgery: the randomised intervention treatment of angina (RITA) trial. *Lancet* 1993;341:573-80.
- 17 McKenna F. Clinical and laboratory assessment of outcome in rheumatoid arthritis. *British Journal of Epidemiology* 1988;27(suppl 1):1-4.
- 18 Felson DT, Meenan RF. Defining clinical improvement in rheumatoid arthritis. *Arthritis Rheum* 1991;34:499-500.
- 19 Silber JH, Williams SV, Krakauer H, Schwartz JS. Using clinical variables to estimate the risk of patient mortality. *Med Care* 1992;30:615-27.
- 20 Bradbury RC, Stearns FE, Steen PM. Interhospital variations in admission severity-adjusted hospital mortality and morbidity. *Health Services Research* 1991;26:407-24.
- 21 Roos NP. Using administrative data from Manitoba, Canada, to study treatment outcomes: developing control groups and adjusting for case severity. *Soc Sci Med* 1989;28:109-13.
- 22 Bion J. Outcomes in intensive care. *BMJ* 1993;307:953-4.
- 23 Iezzoni LI, Daly J. A description and clinical assessment of the computerised severity index. *QRB* 1992;18:44-52.
- 24 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13:818-29.
- 25 Blumberg MS. Biased estimates of expected acute myocardial infarction mortality using MedisGroups admission severity groups. *JAMA* 1992;265:2965-70.
- 26 Hornbrook MC. Hospital case mix: its definition, measurement and use: part II. Review of alternative measures. *Medical Care Review* 1982;39:73-123.
- 27 Iezzoni LI. Severity of illness measures: comments and caveats. *Med Care* 1990;28:757-61.
- 28 Iezzoni LI, Schwartz M, Restuccia J. The role of severity information in health policy debates: a survey of state and regional concerns. *Inquiry* 1991;28:117-28.

(Accepted 14 January 1994)

Equity in the NHS

Equity in community care

Linda Challis, Melanie Henwood

This is the fourth of a series of five articles

School of Social Sciences,
University of Bath, Bath
BA2 7AY
Linda Challis, senior lecturer
in social policy

88 Vicarage Road, Pitstone,
Leighton Buzzard LU7 9EY
Melanie Henwood,
independent social policy
analyst

Correspondence to:
Linda Challis.

BMJ 1994;308:1496-9

The implementation of the NHS and Community Care Act 1990 made local authority social services departments responsible for the organisation and funding of support and care in the community. This development took effect at the same time as a blurring of the boundaries between health and social care. One consequence is that the relevance of equity (a guiding principle of the 1946 National Health Service Act, but relatively lacking from the 1948 National Assistance Act, the foundation of many social services) has come to be more keenly appreciated within personal social services. Equity questions arise in community care over the distribution of public resources between different client groups, income groups, generations, and localities. Moreover, no mechanisms exist to monitor the trends that emerge from different ways that people get access to care. Yet there is a risk that substantial divisive consequences may occur, particularly between generations.

The implementation of the 1990 National Health Service and Community Care Act on 1 April 1993 made local authority social services departments responsible for organising and funding support and care in the community to "enable people affected by ageing or disability to live as independently as possible."¹ Both

the concept of community care and responsibility for its organisation have been notoriously difficult to pin down. For example, a study by the Department of Health and Social Security in 1981 noted the conflicting interpretations of community care by health and social services authorities.² For the NHS, community care usually referred to care provided outside the health service, including local authority residential care. For social services departments it referred principally to non-residential care. The recognition in the 1990 legislation that in practice "people frequently need both social care and health care" is an important acknowledgment of the complexity of this policy area and of evolving distinctions between health and social care responsibilities.

The 1990 legislation was not concerned with the other major users of social services—namely, children and their families. These were the focus of the Children Act 1989, which set out the principles underpinning a new philosophy of partnership between parents and the statutory authorities in the care and protection of children. Unlike the Community Care Act, the Children Act did not, however, introduce a new system for assessment or for funding child care. None the less, local authority social services departments are currently implementing two major pieces of legislation focused on different client groups. Each may have substantial