HOW TO DO IT

Measuring and improving the quality of care

Nicolas U Weir

Stroke Fellow, Calgary Stroke Programme, Room 1211, Foothills Medical Centre, 1403–29th Street NW, Calgary, T2N 2T9, Canada; E-mail: nic.weir@calgaryhealthregion.ca
Practical Neurology, 2004, 4, 340–349

INTRODUCTION

Quality measurement and improvement have an image problem. It just isn't sexy. This is not to say that clinicians aren't interested in providing excellent care, quite the reverse. Nonetheless, compared with treating patients or research, reflecting on daily practice can seem rather mundane. The lack of time and resources traditionally devoted to the topic, together with a recent emphasis on publishing performance 'league tables', have also hampered our proper engagement. In truth, an element of complacency has also held us back. Audit and quality improvement are often regarded as a chore, a box to be ticked or a task to be delegated to junior staff. Inevitably, the fruits of these labours are often superficial, irrelevant, disconnected or even threatening, and contribute to the disenchantment.

Most hospitals still lack effective systems for monitoring and improving the quality of their

clinical services. However, this state of affairs must, and is, coming to an end. Society and its expectations have changed. Patients and public are no longer prepared to trust that they will receive high quality medical care and instead now demand some evidence that this will be the case. The recent finding that *every year* about 850 000 patients admitted to hospital in the UK are harmed by an adverse event requires a response (Department of Health 2000). Unsafe and low quality care is expensive, for example, hospital-acquired infections are estimated to cost the UK National Health Service £1 billion per year and yet 15% may be avoidable (Department of Health 2000). Evidence of involvement in quality improvement is now also a requirement for personal re-validation and hospital re-accreditation in the UK and elsewhere. Clearly, it is time to take the quality improvement agenda seriously.

THE SIX ASPECTS OF QUALITY OF CARE

Effectiveness

Is the best technical care being provided as judged by those with the requisite knowledge and experience?

Acceptability

Is the health care provided in a socially acceptable manner taking into account the views of patients and interested third parties?

Efficiency

Is the provision of care value for money?

Access

Are there barriers between patients and services, such as distance, ability to pay, waiting lists, waiting times or poor supply?

Equity

Are patients treated fairly in relation to others?

Relevance

Is the overall provision of services in line with the health care needs and desires of the population as a whole?

GETTING STARTED ON QUALITY ASSESSMENT

The delivery of health care is complex and a service may be considered excellent, or otherwise, in a number of different ways (see box). Different observers tend to be interested in different aspects of quality. Governments and funding bodies focus on efficiency, equity and relevance whilst patients, public and clinicians focus more on acceptability, accessibility and effectiveness. Understandably, clinicians are most interested in quality of care in terms of its effectiveness and safety, the aspects that best reflect their *raison d'être* and the areas over which they have most control.

Performance within each aspect of care can be measured using the well-known triad of structure, process and outcome. Structure refers to the setting of care in terms of its resources (physical and human) and methods of organization. Process refers to what is actually done and encompasses assessment, investigation, diagnosis and treatment. Outcome refers to any change in the health status of the patient following their care.

Much of the literature over the last 20 years has been devoted to establishing which one of these three approaches provides the best meas-

ure of service quality. To put this into context, it is important to understand that the same period has witnessed government and private organizations introduce, champion, and generally force upon the medical establishment, routine and public systems for measuring *outcome*. Not surprisingly, there has been heated debate.

MEASURING QUALITY USING OUTCOMES

It is easy to understand why outcomes have seemed so attractive (Mannion & Davies 2002). After all, the outcome of disease is what patients care about most, and improving outcome is the fundamental purpose of health care. As such, outcomes have considerable face validity. Furthermore, outcomes are easy to understand and can act as a clinical 'bottom line', summing up the impact of complex packages of care or the skill with which a procedure has been carried out. Where knowledge about the value of treatments is limited or evidence difficult to generalize, it is arguable that measuring outcome is a useful alternative to measuring nothing. Because most hospitals routinely collect information about patient admissions, diagnoses, characteristics and outcome (mostly in terms

$$\text{Outcome}_{\text{overall}} = V_{\text{care}} + V_{\text{measurement}} + V_{\text{casemix}} + V_{\text{chance}}$$

Figure 1 The outcome equation. V, variation.

of case fatality), a cheap and unobtrusive data collection system would also already appear to exist.

However, because medical care is only one of many factors that influence the outcome of disease, the use of outcome as a quality indicator is problematic. In particular, when faced with variation in outcome, it is difficult to know whether this represents the 'signal' of a real difference in quality of care or simply the 'noise' generated by measurement error, differences in the types of patient treated, or the play of chance (Fig. 1). Unfortunately, in most cases, the signal is rather small. Even in a randomised trial, most interventions 'that work' have only a modest impact on outcome. In everyday practice, where the patients may be less 'ideal' and the differences in the proportions receiving the intervention are smaller, the impact of any differences in care is likely to be smaller still. Consider, for example, two hospitals, each with a 30 day case fatality of 25% for patients with an acute stroke admitted to an ordinary ward. Even if one hospital admitted 50% to a stroke unit [leading to a 14% relative reduction in case fatality for them (Stroke Unit Trialists' Collaboration 2004)] and the other hospital none, the absolute difference in case fatality between the two hospitals would only be 2% (23% vs. 25%). Can the impact of bias, confounding and the play of chance really be reduced to thresholds as low as this when measurements are made in the real world? Perhaps not surprisingly, for most medical conditions, the answer is no, and it is worth understanding why this is so.

The problem of what to measure, and when

Often death is used as an outcome indicator and, for life-threatening disorders, this is understandable. However, most patients with serious disorders do not die but many fail to make a complete recovery (e.g. stroke or myasthenia gravis), whilst other patients have less serious conditions that nonetheless have a major impact on their health and well-being (e.g. multiple sclerosis or migraine). To measure the impact of treatment in those who don't die, it is necessary to collect data that describe their quality of

survival, i.e. symptoms, impairments, activities, participation, quality of life, etc. Unfortunately, the great strength of these data – their qualitative nature – is also their Achilles heel in that they are intrinsically more prone to measurement error and, because they are not routinely reported, they are more difficult and costly to collect. This is particularly true for the outcomes of greatest importance to patients (activity, participation and quality of life), where the scope for subjectivity and confounding is greatest (Duncan *et al.* 2000). Confidently linking these outcomes to the quality of care provided is therefore difficult. If routinely reported variables such as discharge destination, readmission or complication rates are measured instead, the problems of interpretation only increase. For example, whether a patient with traumatic brain injury is discharged to their own home or to residential care depends not only on their clinical state but also on the wishes and capabilities of the family, their financial and housing arrangements, and local cultural norms and social supports, over which a hospital has no control.

The use of hospital discharge data leads to other errors. In particular, hospital samples may be biased by local diagnostic and coding habits. A hospital might appear to have better stroke outcomes than is truly the case simply because its physicians diagnose transient ischaemic attack as minor stroke more often than others, or because its coding staff more often code stroke as a secondary diagnosis (for example, if the patient dies from a complication such as pneumonia). Similarly, because it is convenient, outcomes are often measured at the time of discharge rather than at a defined time after symptom onset. This too may mislead. For example, in Cleveland, USA, a large scale outcomes measurement system showed year on year reductions in-hospital case fatality for a number of conditions, leading many to conclude the system had led to an improvement in quality of care (Baker *et al.* 2002). However, later analysis showed that 30 day case fatality had actually stayed the same, the deaths simply having been shifted to the post-discharge period as patients were discharged earlier in the later years.

Figure 2 Hypothetical distributions of case fatality at hospitals with acceptable quality of care; hospitals with poor quality of care; and all hospitals combined, assuming perfect adjustment for casemix. A large proportion of hospitals with outcomes above the outlier cut point have acceptable quality of care whilst a large proportion of hospitals with poor quality of care remain in the body of the outcomes distribution.

‘Measuring quality of care using outcomes is like looking for bad apples – it tells you nothing about most of the apples’.

The problem of comparing like with like – casemix

To ensure like is being compared with like, statistical techniques are often used to adjust comparisons of outcome for differences in the types of patient treated (i.e. differences in casemix). However, these adjustments are fundamentally limited by our incomplete knowledge of *all* the factors that influence outcome. Even under ideal circumstances, statistical models usually explain only a proportion of the total variation in outcome. Under less ideal circumstances, such as when routinely reported data are used to adjust for casemix, this problem is compounded by the reduced availability and relevance of predictive factors and by error in their measurement. Indeed, the impact of inaccurate measurement of predictive factors rivals the impact of differences in quality of care on comparisons of adjusted outcome (Green & Wintfeld 1993). Even bespoke systems for collecting predictive data may be incomplete and unreliable (Fine *et al.* 2003). Where there is no single and accepted predictive model, conclusions regarding hospital performance vary depending on the model that is used to adjust for casemix (Iezzoni 1997). Thus, despite 'adjustment', it is rarely safe to assume that like really is being compared with like.

The problem of chance

Chance has a much greater impact on outcomes than is usually appreciated. Firstly there is the simple question of numbers. Even for common medical conditions, most hospitals admit relatively few patients per year (usually a few hundred) of which fewer still experience an outcome (such as death). In any one year therefore there is considerable impression in the estimates of outcome, limiting both the ability to differentiate between hospitals and making changes in ranking over time more likely to reflect regression to the mean than changes in the quality of care.

Secondly, hospitals are conventionally identified as having a potential quality problem if their outcome falls within the worst 2.5% of the outcome distribution (a 'quality outlier'). However, by chance alone, a proportion of hospitals with good quality care will appear as outliers and a proportion of hospitals with poor quality care will have outcomes that remain within the 'acceptable' body of the outcome distribution. This is best appreciated graphically (Fig. 2). In fact, simulation studies suggest that even under ideal conditions the positive predictive value and sen-

sitivity of outcomes data for common medical conditions are surprisingly low. For example, with a sample size of 900, the positive predictive value of outlier status may reach only 32% and the sensitivity only 68% (Thomas & Hofer 1999). Moreover, at many hospitals, samples of this size can only be achieved by collecting data over several years, reducing the usefulness of the data.

Bad apples and poor signposts

To all this can be added that, when used to make comparisons, outcomes inherently divide providers into those either likely or unlikely to have deficiencies in care – the so-called hunt for 'bad apples'. This approach ignores the fact that every provider *always* has the scope to improve their services in one way or another. By allowing the majority to shelter in the body of the outcomes distribution, opportunities for improvement are obscured and an attitude of complacency fostered. And even if outcomes do correctly identify a provider as having poor quality of care, they do not identify the remedy. Outcomes are simply a prelude to further investigation, not a signpost to action.

Useful outcome measurement

For most medical conditions therefore outcomes are simply not suited to the task of measuring and improving the quality of care. However, this is not to say that outcomes never have a role in quality improvement:

- they can act as useful quality indicators in some areas of elective surgery (where the attribution of outcome is easier);
- they may have a role in 'in-house' monitoring systems, e.g. infection rates;
- individual adverse outcomes may act as useful catalysts for an in-depth service review;
- and, in all settings, outcomes provide useful contextual information.

However, to properly understand the quality of a service, most effort should be focused on measuring the structure and process of care itself.

MEASURING QUALITY BY STRUCTURE OF CARE

As the stage on which the processes of care are played out, structural items are necessary but not sufficient for good medical care. The collection of structural information is therefore a blunt way to measure quality. This seems particularly true for simple counts of staff numbers, their qualifications and the availability of equipment.

For example, it is inconceivable that a high-quality stroke service could function without access to a CT scanner, but the mere presence of a scanner does not mean that patients will be scanned in a timely manner. However, there is growing interest in structural items that refer to how clinicians organize themselves and communicate, stemming from evidence that better practice in these areas may directly improve care and outcomes. For example:

- organizing stroke care in a geographical unit and instituting regular, goal-orientated team meetings is known to result in better processes of care and outcomes (Stroke Unit Trialists' Collaboration 2004);
- concentrating high-risk surgery or paediatric intensive care in centres of excellence can be shown to reduce mortality (Dimick *et al.* 2003; Tilford *et al.* 2000);
- computerized reminder systems or nursing protocols improve uptake of preventive care in hospitals (Dexter *et al.* 2001; Rhew *et al.* 1999).

Data such as these are relatively easy to collect and might form realistic quality improvement targets. Nonetheless, if the goal is to measure the effectiveness of care, then, wherever possible, the collection of structural information should be seen as an adjunct to the main effort – measuring the *process* of care.

MEASURING QUALITY BY PROCESS OF CARE

In comparison with outcomes, process data have a number of important advantages (Mannion & Davies 2002; Crombie & Davies 1998). Process data are:

- More easily measured. Compare, for example, the difficulty in measuring functional outcome after stroke with the ease of using medical records or discharge data to see whether, and for how long, a patient was admitted to a stroke unit (the main arbiter of a better functional outcome under hospital control).
- More easily interpreted. Process data more directly reflect the care that was provided and are much less influenced by variation in casemix. They can also be interpreted in relation to research evidence, avoiding the need to make comparisons between hospitals.
- More sensitive. Where comparisons are made, smaller sample sizes are required to arrive at confident conclusions. For example, it has been estimated that to confidently draw attention to a 49% uptake of important in-

terventions for acute myocardial infarction at one hospital vs. zero uptake at another would require a sample of 495 patients using outcomes data but only 15 using process data (Mant & Hicks 1995). Process data therefore allow more timely measurements of quality.

- An indicator for action. Process data suggest the actions needed to bring about improvement. For example, if only 60% of patients with an ischaemic stroke are discharged on aspirin, attention is immediately focused on remedying this fact.
- Universally useful. Process data highlight areas for improvement in virtually every service under investigation. The approach therefore helps to 'move the mean' of quality of care across all hospitals to the right.

In addition, process data provide:

- the only means to monitor and prevent 'near misses' (potentially serious deficiencies in care that uncommonly lead to adverse outcomes);
- lessons from otherwise overlooked outcomes (e.g. the death of an elderly patient with a severe stroke may still have been contributed to by poor quality of care);
- information on unnecessary resource use.

If structure and process data are to be used, the only stipulations are that the measurements should address important aspects of care and that better performance in each should be proven or strongly believed to lead to better outcomes. Beyond this, the approach may be either formal or opportunistic.

Formal methods of measuring structure and process

The formal approach refers to 'instruments' that can be used 'off the shelf'. Here, the development of a set of reference criteria usually starts with a systematic review of the literature and grading the research evidence. Existing high quality systematic reviews and clinical guidelines can be used as a template. A panel of experts is usually convened to revise the criteria and to attach acceptable time limits and caveats to deal with variation in casemix. The latter may take the form of explicit extenuating circumstances [e.g. the failure to hold a multidisciplinary team (MDT) meeting should not be penalized if a patient with a minor stroke is discharged within 48 hours of admission]. The Intercollegiate Stroke Audit Package is an excellent example of this approach (Gompertz *et al.* 2001). It surveys key items of the organization and process across the spectrum of stroke care (physicians, nurses,

therapists; admission through to discharge) using data collected from medical records.

Opportunistic methods of measuring structure and process

The opportunistic approach is particularly attuned to the nitty-gritty of systems of care and issues of patient safety 'in-house'. The approach is fluid and the starting point usually a particular area of concern, highlighted perhaps by a recent discussion or clinical incident (Cook *et al.* 2004). For example, the diagnosis of yet another case of pneumonia in a patient with stroke may spark interest in the quality of the systems for its prevention. The next step is an attempt to understand the size of the problem, for example, by surveying the numbers and types of patient who develop pneumonia and the adequacy of preventive interventions. Next, the reasons for any shortcomings are explored, usually through further observation, interview and discussion. Continuing the example, it may transpire that admitting staff are unaware of the possibility of silent aspiration in patients with seemingly mild stroke; that swallow assessments are being carried out inappropriately or with undue delay; that methods of hydration are inadequate; or that mobilization is limited by staff or equipment shortages. With this information in place, a rational plan for change can then be implemented.

Difficulties with structure and process reviews

Measurements of structure and process have certain limitations (Mannion & Davies 2002). In particular, in many areas of medicine, firm evidence of what constitutes effective care is still limited. As a result, staunchly evidence-based reviews may lack content validity whilst more comprehensive reviews may end up measuring many items of uncertain importance. Even 'firm' evidence is sometimes difficult to generalize to daily practice. For example, whilst the concept of stroke unit care is of proven effectiveness, its effective components are not well understood, and simple comparisons of the proportion of patients admitted to 'a stroke unit' may hide important differences in service quality. That said, such shortcomings can be minimized. For example, comparisons of stroke unit care are less likely to mislead if we also survey key aspects that we suspect make them important (early admission, early assessment by therapists, etc.).

Other difficulties relate to the fact that process measurement often entails the collection of data from the medical record. Much of what is measured may therefore reflect the completeness of record keeping rather than what was actually done and with how much skill. And explicit criteria sometimes fail to allow for the subtle factors that guide the decision whether or not to provide treatment. Furthermore, all formal review criteria are prescriptive and so run the risk of stifling clinical innovation and of becoming obsolete. Finally, abstraction of data from medical records is prone to observer error, especially if clinicians audit their own activity with the inevitability of bias in their own favour, and it is time consuming.

Nevertheless, we must be pragmatic. Despite their shortcomings, in most situations, measurements of the structure and process of care are more likely to point us in the right direction than measurements of outcome. For example, the benefits of this approach are borne out by the UK National Sentinel Audit for Stroke, which is helping to realize tangible improvements in stroke care, whilst the positive impact of the routine comparisons of outcome after stroke has been hard to discern (Rudd & Pearson 2002; Mannion & Goddard 2003).

USING PERFORMANCE DATA TO CHANGE PRACTICE

Simply deciding on the tools of measurement is only half the battle. Measurement data are only likely to bring about improvements in care if they are released into a favourable environment. Critical to this are the questions: why and for whom are the data being collected? Are the data chiefly concerned with external accountability (satisfying stakeholders that the service is adequate) or is their purpose to understand what is going on (to inform clinicians about opportunities to improve)? Regrettably, in recent years this distinction has been fuzzy. In particular, routinely collected outcomes data have been made public as 'league tables' with the supposed purpose of informing public and purchasers *and* identifying opportunities for service improvement. The attempt to have matters both ways comes up hard against human nature.

Inevitably, humans are reluctant to be tested, especially if they think the test is unfair and if the results are to be released to a public and media that do not understand the niceties of the data. As a result, clinicians have tended to view pub-

lished performance data as an external threat to be 'dealt with' or ignored rather than as a resource to improve their service. Interestingly, there is little evidence to suggest that stakeholders have made much use of the data either (Mannion & Goddard 2003).

More insidiously, the use of performance data to hold providers to account can have unintended and negative consequences, as humans understandably attempt to avoid being punished. For example, for services judged using outcomes, there may be the temptation to refuse to treat patients likely to have a poor outcome or to quickly discharge them so their poor outcomes are counted elsewhere (see the example from Cleveland Baker *et al.* 2002). And whatever the method of measurement, the figures can be fiddled (gaming), either through dishonest reporting of the care provided or by making patients appear sicker than is truly the case, instantly improving adjusted outcome, as seemed to occur in New York State when it first published league tables for coronary artery by-pass graft surgery for individual surgeons (Green & Wintfeld 1995).

CREATING LEARNING CULTURES

If a system for promoting quality improvement is to be effective, a less-threatening approach must be adopted. Here we must learn from industry that has a long history of successful performance measurement. The key appears to be the creation of a culture in which failure can be openly discussed and the operation of smooth-running systems reviewed; in which the detailed knowledge of front-line staff is recognized as the best source of opportunities for improvement; and in which knowledge for improvement is disseminated widely using formal and informal means (Davies & Lampel 1998). Under-pinning all this is the clear perception that performance measurement is for understanding rather than making judgements.

Learning cultures are decentralized, place considerable emphasis on team work and are fundamentally open and trusting – the very opposite of the centralized, hierarchical and somewhat suspicious structures often found within health services. Movement towards learning cultures will therefore require considerable commitment from all. For the purposes of accountability, governments and purchasers should de-emphasize comparisons of outcome and instead focus on checking that hospitals have the right systems in

place, and are using them, to improve quality of care (Lilford *et al.* 2004). They should also ensure that the necessary time, resources and incentives are made available. As clinicians, we must be prepared to properly engage with and even champion quality improvement. In doing so, we should familiarize ourselves with the skills of group-working, leadership, and bringing about change, skills we tend to lack (Wilcock & Campion-Smith 1998), and with new statistical techniques, such as quality control-charts, to help quickly spot deviations from usual standards of care (Lilford *et al.* 2004). It will be interesting to see whether, and how far, these challenges are met.

NEUROLOGY

As for neurology, the problems we face in measuring quality of care are generally much the same as those that face others. However, a few points can be made. The measurement of outcome is probably more difficult in neurology than in some other specialities. Death is seldom a useful indicator: beyond stroke, most disorders we deal with are rarely fatal in the short term, and in others, such as motor neuron disease, death is inevitable and may even sometimes be desired. Much more often we deal with patients with chronic symptoms and/or progressive disability and the disorders themselves may be uncommon or rare. Obtaining valid measurements of the appropriate qualitative outcomes in these patients is challenging in randomised trials, let alone in ordinary care. And until recently, the measurement of the structure and process of care might have seemed superfluous – are we not supposed to be the speciality that diagnoses in great detail but provides little effective care? Latterly, however, this jibe has started to ring hollow. The number of proven treatments and high quality clinical guidelines is expanding, and meaningful structure and process review is possible for most important neurological disorders. Indeed, an important task for the speciality is to derive methodologically sound review instruments for common disorders along the lines of the Intercollegiate Stroke Audit Package (Gompertz *et al.* 2001). However, beyond this detail, perhaps the most important task for neurologists is to understand that the agenda has changed. Measuring and improving the quality of neurological services is no longer optional. We have a duty to respond constructively.

CONCLUSIONS

- The quality of a system of care can be measured by its structures, processes and outcomes.
- For medical conditions, identifying variations in outcome that can be confidently attributed to variations in the quality of care is very difficult because of the greater impact of measurement error, casemix and the play of chance and because of our limited ability to minimize and allow for this fact.
- By drawing attention to 'bad apples', comparisons of outcome fail to inform about the quality of care in the majority. Reliance on measuring outcomes can foster an attitude of complacency in many cases.
- Structural items are easy to measure but simple counts of personnel and resource give blunt indications of service quality. A focus on methods of organization and communication may be more fruitful.
- Compared with outcomes, process data are easier to measure and interpret, more timely, and they are an indicator for action in nearly all systems of care. Wherever possible, measurements of quality should focus on measurements of the process of care (by formal and informal means).
- The drawbacks of structure and process measurements should be kept in mind. In particular, measurements are difficult when the evidence base is small or difficult to generalize. And they frequently refer to the adequacy of record keeping rather than what was actually done.
- Simply measuring quality is unlikely to bring about improvement. Systems that emphasize public accountability run the risk of unintended and even negative consequences for patient care.
- To properly engage with quality improvement, medicine must adopt the values of learning cultures. Whether governments, purchasers, managers and clinicians respond constructively remains to be seen.

ACKNOWLEDGEMENTS

This article was reviewed by Professor Martin Dennis, Edinburgh.

REFERENCES

- Stroke Unit Trialists' Collaboration. (2004) Organised inpatient (stroke unit) care for stroke. *Cochrane Stroke Group. Cochrane Database of Systematic Reviews*, **4**.
- Baker DW, Einstadter D, Thomas CL *et al.* (2002) Mortality trends during a program that publicly reported hospital performance. *Medical Care*, **40**, 879–90.
- Cook DJ, Montori VM, McMullin JP, Finfer SR & Rocker GM (2004) Improving patients' safety locally: changing clinician behaviour. *Lancet*, **363**, 1224–30.
- Crombie IK & Davies HTO (1998) Beyond health outcomes: the advantages of measuring process. *Journal of Eval Clinical Pract*, **4**, 31–8.
- Davies HTO & Lampel J (1998) Trust in performance indicators? *Quality in Health Care*, **7**, 159–62.
- Department of Health. (2000) *An Organisation with Memory: report of an expert group on learning from adverse events in the NHS*. The Stationary Office, London.
- Dexter PR, Perkins S, Overhage JM *et al.* (2001) A computerized reminder system to increase the use of preventive care for hospitalized patients. *New England Journal of Medicine*, **345**, 965–70.
- Dimick JB, Pronovost PJ, Cowan JA Jr, *et al.* (2003) Variation in postoperative complication rates after high-risk surgery in the United States. *Surgery*, **134**, 534–40.
- Duncan PW, Jorgensen HS & Wade DT (2000) Outcome measures in acute stroke trials. A systematic review and some recommendations to improve practice. *Stroke*, **31**, 1429–38.
- Fine LG, Keogh BE, Cretin S, Orlando M & Gould MM, for the Nuffield-Rand Cardiac Surgery Demonstration Project Group (2003) How to evaluate and improve the quality and credibility of an outcomes database: validation and feedback study on the UK Cardiac Surgery Experience. *British Medical Journal*, **326**, 25–8.
- Gompertz PH, Irwin P, Morris R *et al.* (2001) Reliability and validity of the Intercollegiate Stroke Audit Package. *Evaluation in Clinical Practice*, **7**, 1–11.
- Green J & Wintfeld N (1993) How accurate are hospital discharge data for evaluating effectiveness of care? *Medical Care*, **31**, 719–31.
- Green J & Wintfeld N (1995) Report cards on cardiac surgeons. Assessing New York State's approach. *New England Journal of Medicine*, **332**, 1229–32.
- Iezzoni LI (1997) The risks of risk adjustment. *Journal of the American Medical Association*, **278**, 1600–7.
- Lilford R, Mohammed MA, Spiegelhalter D & Thomson R (2004) Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet*, **363**, 1147–54.
- Mannion R & Davies HTO (2002) Reporting health care performance. learning from the past, prospects for the future. *Evaluation in Clinical Practice*, **8**, 215–28.
- Mannion R & Goddard M (2003) Public disclosure of comparative clinical performance data: lessons from the Scottish experience. *Evaluation in Clinical Practice*, **9**, 277–86.
- Mant J & Hicks N (1995) Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *British Medical Journal*, **311**, 793–6.
- Rhew DC, Glassman PA & Goetz MB (1999) Improving pneumococcal vaccine rates. Nurse protocols versus clinical reminders. *Journal of General Internal Medicine*, **14**, 351–6.
- Rudd AG & Pearson M (2002) National stroke audit. *Clinical Medicine*, **2**, 496–8.
- Thomas JW & Hofer TP (1999) Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. *Medical Care*, **37**, 83–92.
- Tilford JM, Simpson PM, Green JW, Lensing S & Fiser DH (2000) Volume–outcome relationships in pediatric intensive care units. *Pediatrics*, **106**, 289–94.
- Wilcock PM & Campion-Smith C (1998) Never mind the quality: feel the improvement. *Quality in Health Care*, **7**, 181.