Audit, assessment or performance management?

Mike Pearson and Jane Ingham

Mike Pearson
FRCP, Director

Jane Ingham
Manager

Clinical Effectiveness and Evaluation Unit, Royal College of Physicians

Clin Med 2005;**5**:7–11

When did anyone last check on your clinical work? Consultant physicians make many decisions each day. Each decision about an individual patient has to last hours, days or weeks, and many will have profound effects (most, but not all, beneficial). Prior to appointment, each consultant has passed difficult exams to ensure adequate knowledge, and been supervised and monitored in training. But once appointed, direct assessment ceases and clinicians are free to practise almost unhindered. Only if there is a complaint or a clinical disaster does anyone ask questions or inspect the case records, and even then it is unusual to delve into how a consultant performs behind the closed consulting room door.

Clinical audit was proposed 20 years ago as a means of assessing objectively whether good care was being delivered, but has never really lived up to its promise. Audit studies in local settings have all too often been performed on small numbers of cases, with unvalidated questionnaires or other assessments, and without enough rigour to establish whether the observations were meaningful. Even when problems were noted, there was often no system within the hospital that could lead to change. The Bristol Enquiry[1] noted that the NHS was awash with data, but most was unused. However, physicians mostly care for chronic conditions where there are few existing data, few measurement tools, and even fewer systems to do anything about it.

If the public are to be assured that medical care really is as good as we believe it to be, then there must be an understandable and reportable measurement. This requires that there are agreed standards, demonstrably reliable means of measurement, reports that are clinically relevant, and that all clinicians have taken part. This is the challenge that the Clinical Effectiveness and Evaluation Unit (CEEu) has been trying to tackle since its inception seven years ago.

Fundamentals of obtaining information

It is much easier to talk about obtaining and using data than it is to collect it. In his address at the opening of the RCP Information Laboratory (*i*Lab) last year, Derek Wanless observed how difficult it had been to set up reliable systems in the banking sector – and that in the NHS we had to consider not only numbers but the human factors associated with the abilities and operating environment of local medical teams on the one hand, and the rights of patients to be individuals on the other. Adding unpredictable human factors to the evaluation makes the problem much more challenging. The next three sections will discuss the hurdles that must be, and can be, overcome.

Choosing and setting standards

Medical science is not exact. Standards applicable to most patients have emerged via guidelines, via government initiatives and via *ad hoc* projects. Their origins affect their usefulness.

Guidelines. National guidelines have proliferated since the first UK national guideline appeared in 1990.[2] Over 250 current guidelines for physician medicine are listed on the RCP guideline database.[3] Most of them have emanated from specialist societies, and a few from the national collaborating centres (NCCs) sponsored by the National Institute for Clinical Excellence (NICE). The NCC products have extremely high standards of evidence-based medicine but at a production cost well beyond the capacity of voluntary societies. However, most NICE-commissioned guidelines on chronic conditions have concentrated on clinical issues and avoided comment on the service implications, ie the context in which clinical activity occurs. Yet it is impossible to deliver clinical care without considering clinical and service aspects together. Most of the guideline evidence derives from studies about the use of drugs and interventions (because the pharmaceutical industry has sponsored such studies to get their products registered), rather than on the reliability of diagnosis, or on how to assess clinical outcome. Not surprisingly, therefore, even our NICE guidelines, with their emphasis on assessing all the evidence, contain a majority of Grade D recommendations, which are based on expert consensus in the absence of evidence.[4] Nevertheless, the presence of a nationally accepted guideline is an important starting point when setting out to measure good practice.

Government initiatives. The Department of Health has set numerous standards and targets over the years, many founded on little evidence. However, the national service frameworks (NSFs) have considered

topics in depth and in particular how to set targets that will challenge the clinical professions to reach better standards that should lead to better patient outcomes. The NSFs specifically consider the service implications and use the evidence to support their targets. Thus the cardiology target for 'thrombolysis within 30 minutes of arrival at hospital' is based on clinical trial data but is a measure of how well a hospital team is coordinating the care processes that culminate in thrombolysis; ie it is a marker of the overall effectiveness of a team of healthcare professionals performing several different functions.

Ad hoc *approaches.* When there is no recognised guideline or NSF, it is still possible to examine the literature and assemble expert groups to define meaningful targets. Thus, for lung cancer, the RCP brought together a group to define what 'needed to be known' to understand the processes and outcomes of care, and within six months had published a document supported by all the major clinical groups.[5] This allows a notional benchmarking standard to be derived. It demonstrates what is possible, although sadly other Department of Health activities led to the initiative being marginalised for three years and only now is the audit led by CEEu beginning its national roll-out.

Measurement

Few clinical data are routinely collected. NHS IT systems record the quantity, rather than the quality, of care delivered. In the early days of clinical audit, many stressed that audit was not research. This was true in that audit was not intended to reveal new knowledge, but it was unhelpful since it encouraged local auditors to perform small studies. Their observations could have arisen by chance so it was not justifiable to alter practice based on the data. The processes of 'evidence-based medicine' insist that papers are analysed according to explicit standards intended to reduce bias or chance. Thus future care delivery is influenced by medical 'facts'. Those producing assessments of clinical performance should expect to work to similar standards. There are three linked stages: define the question; ensure reliability; and report intelligently. Each sounds simple but the collaborative work required to achieve them is considerable, especially if the processes are to be widely accepted as fair and appropriate.

1 Each audit project must begin by defining what it wants to know about the condition and then limiting the variables to only those needed to answer the particular question. This means thinking ahead to how the data are going to be reported back to clinical and managerial teams and making sure that there is enough to answer the question and to control for major confounding factors.

2 The Stroke Audit's initial data collection in 1988 was planned by an expert multidisciplinary group and refined with Delphi surveys. Yet when the pilot study was performed and the same variables collected from the records by two different auditors, the inter-observer reproducibility was little better than chance. The project team had to revisit the variables and their definitions, and then re-write the help leaflets for auditors. Only then did duplicate data collection confirm that the results were robust.[6] Experience in audits on chronic obstructive pulmonary disease (COPD) and continence has emphasised the need for this attention to detail – and perhaps is a reason why many local audits have struggled to produce results worthy of the time expended on the data collection.

   There are two methods of data collection in audit projects: snapshot and continuous data collection.

   - The snapshot approach collects data on a set number of cases per hospital simultaneously. It can collect much more detail per case, and can report on how hospital facilities compare to care delivery as well as providing inter-hospital comparisons. Because only a limited number of cases are collected, the workload for data collection staff locally is finite. Reports can usually be completed within 10–12 weeks of data being received.
   - The continuous audit collects data on a more limited set of variables but on every case from each hospital in an ongoing manner. There is a continuing workload for local hospitals that is often not built into resource allocations, but the benefit has been that with electronic data collection reports can be made available within 24 hours and it is possible to build up a longitudinal picture of care delivery.

3 The way results are reported matters: the data must be presented in ways that are easily understandable by an audience that may include medical and non-medical professionals, lay managers and patient groups.

Recruitment

A unique feature of the recent RCP audits has been the extent of the coverage. Audits led by single specialty societies have struggled to obtain data from more than 25% of hospitals. By contrast, the Stroke Audit and the Myocardial Infarction National Audit Project (MINAP) now have participation from all hospitals in England and Wales. The first Stroke Audit in 1998 recruited just over 80% of hospitals and in 2004 had participation from 100% of hospitals in England, Wales and Northern Ireland. MINAP recruited 90% of trusts over about 18 months and the last 10% in the next 18 months. The COPD Audit achieved 94% participation across the UK, and the Blood Transfusion Audit 71% of trusts. The Continence and Lung Cancer Audits are about to roll out nationally and are expected to match these participation levels. It is not just that some audits are associated with NSFs or national targets – the COPD Audit has no associated initiative. The RCP name appears to carry greater 'weight' than individual societies, and the team within the RCP has been single minded in persuading hospitals to take part. However, recruitment is not easy – few volunteer for extra work. There has to be something for the trust in return, and that has to be confidence that the coordinators will deliver useful data back to the trust.

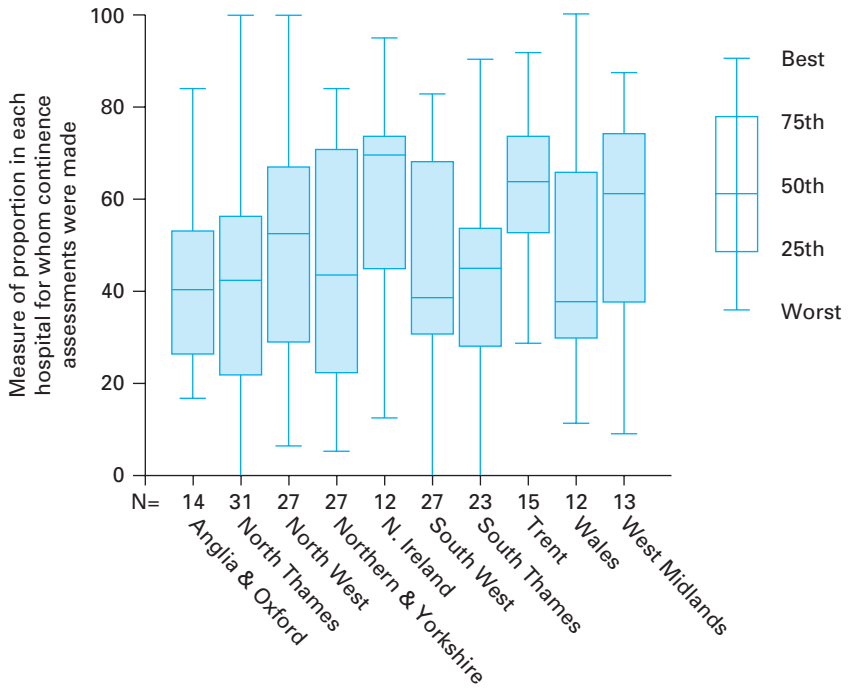

Fig 1. Box plot to illustrate the best, 75th, 50th, 25th and lowest performance of the continence indicators in each trust within each region of the country. From the first Stroke Audit.

Results

Results will be illustrated with reference to two projects, the Stroke Audit and MINAP.

The Stroke Audit is in its fourth iteration using the snapshot approach. After each round of audit each hospital received a report showing how they fared in comparison with the national (and sometimes regional) picture. In many cases, results have been compared between hospitals using 'box and whisker' plots to illustrate the variability. Figure 1 comes from the first Stroke Audit and demonstrates the variability between hospitals on continence care after stroke. About half of stroke patients have continence problems, yet in every region of the country some hospitals check on nearly every patient and others on none. Expressing data in this form avoids league tables but those at the lowest end of performance spectrum know it is possible to do better since their colleagues are doing just that. But there is an important caveat. Apparent poor performance may be due to inadequate data collection, lack of facilities, or sometimes just unrepresentative cases. It should therefore be an indication for investigation but not for automatic castigation. Also, as the range of performance narrows, so the statistical power to discriminate between hospitals is reduced.

The stroke data have been used very actively by trusts, and most improvements are probably the result of local responses. However, data are now shared with the strategic health authorities (SHAs) and the Healthcare Commission (formerly the Commission for Health Improvement), and inform the inspections of trusts. In addition the data have informed the guideline development process and the structures of the NSF.

Associated with this intense activity, sequential audits have shown significant improvements in stroke care delivery, including a reduction in hospital death rates with no change in levels of residual disability at discharge (Table 1). It is impossible to determine with any certainty how much the audit data have been causal, but the availability of data that are accepted and respected has been part of the story.

MINAP was fortunate to link to a secure, efficient and flexible database at the Central Cardiac Audit Database.[7] From this foundation, the MINAP team were able to assure confidentiality of data and to recruit hospitals to take part. Data reports were available to the participants within 24 hours and quarterly reports were made available to trust managers and to SHAs. As confidence has grown, results from individual hospitals have been shared with both the public and the Healthcare Commission, which incorporates some results within their star

Table 1. Comparative data from the first and third rounds of Stroke Audit data (1998 to 2002) showing improvements in care over that period.

	1998	2002
% Discharged alive	65	70
Length of stay (survivors)	35 days	39 days
% Hospitals with stroke units	49	77
% Hospitals with stroke specialists	65	83
% >50% time on stroke unit	16	28
% Visual field checked	44	65
% Urgent scan	51	57
% Seen by OT within 7 days	44	51

OT = occupational therapist.

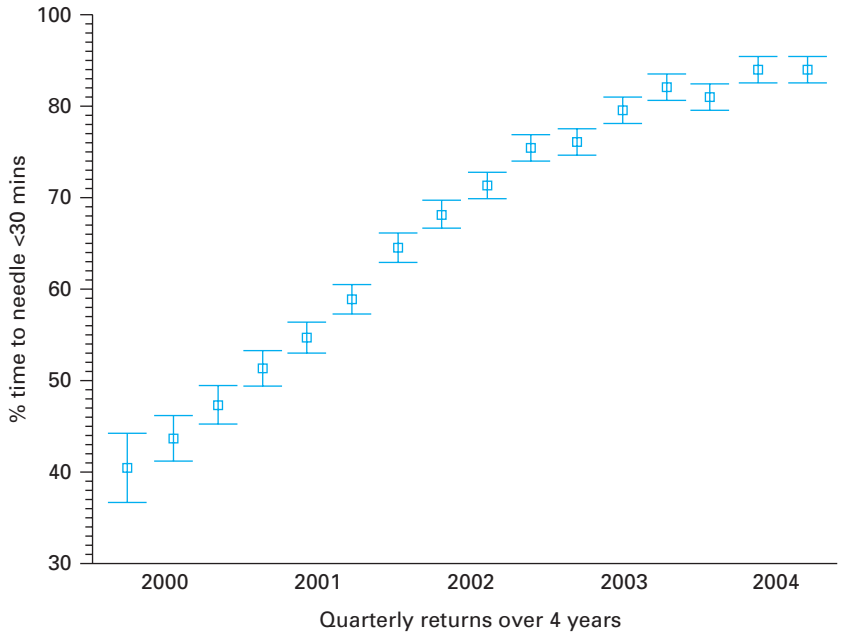

Fig 2. The proportion of patients with an ST elevation myocardial infarction receiving thrombolysis within 30 minutes of arrival at hospital over three years. From the MINAP study.

ratings system. There have been other NSF-related cardiology initiatives running in parallel with MINAP, but without reliable measurements these initiatives would not have known which hospitals were or were not making progress toward the target. Nor would local units have had confidence that the requests made of them were appropriate or achievable. The performance of trusts in meeting the door-to-needle and the secondary prevention targets is little short of astonishing (Fig 2), especially as this has occurred at a time when A&E departments have been under much stress. There have been associated falls in the 30-day mortality of ST elevation infarcts in line with that predicted by the randomised controlled trials (RCTs) of 15 years ago.[8]

Project structure

Over the past seven years, the CEEu has developed considerable expertise in running audit projects, and has benefited from running projects in parallel, as lessons from one are passed on to the next. The success of the CEEu in achieving national coverage in three topics with two more on the way (and others mooted) is without parallel. The Intensive Care National Audit and Research Centre (ICNARC)[9] audit of intensive care with 74% coverage is probably the closest challenger. There are a number of features that have contributed:

- a 'critical mass' of researchers that cannot be attained in smaller organisations
- the RCP name (linked to a specialist society and other partners), which has provided confidence in the project, and was important in attracting funding
- multidisciplinary partnerships (including patients) that oversee the development and interpretation of the data and results
- reporting on data in ways that are perceived to be helpful by clinicians, managers and others
- project management that consistently delivers products on time and within budget (time lines that have been promised must be met).

Of these, perhaps the most important has been the firm management of each project so that the collaborators in hospitals across the country get what they expect. The reputation of most government-funded projects is for them to 'drift' and either deliver late or not at all. These projects have kept to their time lines and the staff involved deserve particular commendation for making this happen.

The future

These projects have moved clinical audit forwards into a form of comparative clinical performance assessment. They operate to research standards and are probably better termed 'health service research', since not only do they produce data on performance but they also yield much useful observational/epidemiological data on methods of service delivery. As an example, the MINAP dataset links to the Office of National Statistics and has been able to demonstrate marked differences at one year between those sent home with and those sent home without a statin drug (Fig 3). It is helpful to confirm that the rather selective nature of RCT studies does translate into real-life clinical practice although these are early uncontrolled observations and do not yet prove the point.

When the National Program for IT is able to collect more clinical data, such longitudinal monitoring may be possible for many conditions. But for the moment that is only a tantalising possibility and audit remains dependent on specific data collections.

These projects do have a cost: local costs of collecting the data

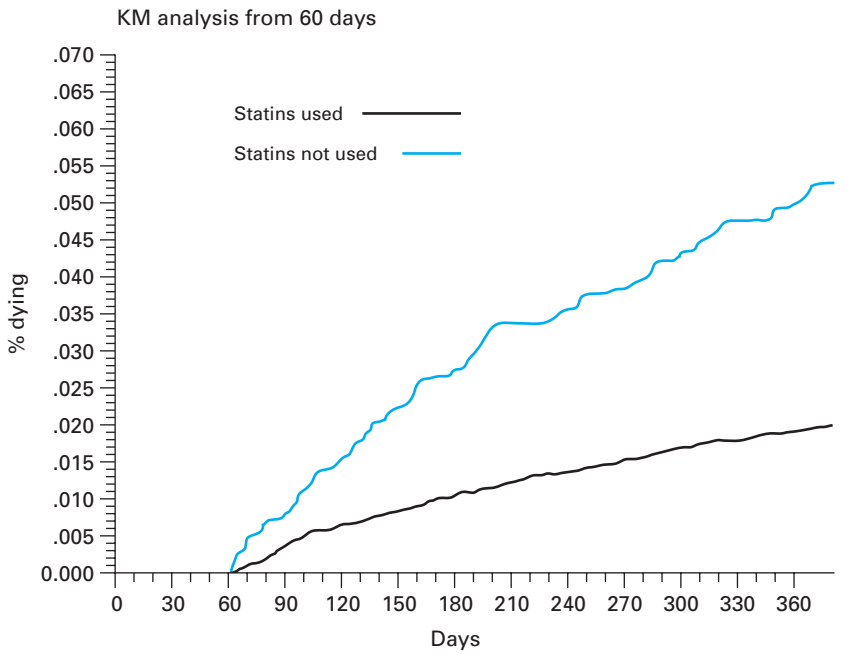

Fig 3. Proportion of patients who die in the period 60–360 days post discharge according to whether they did or did not go home with a prescription for statin drugs. The first 60 days are ignored since various active interventions in that time may skew the data. From the MINAP study.

and central costs of analysis and reporting. There are three ways of justifying the expenditure:

1 Most industrial companies expend between 3 and 5% on quality assurance and, given that healthcare outcomes are so important to customers, perhaps the NHS should do the same. The local and central NHS cost of collecting a case for MINAP is about £15 compared to £1,500 or more per admission, ie 1% or less of the cost of hospitalisation. The stroke snapshot approach works out at even less.

2 It would be ideal to compare costs with benefits, but these audit projects are not randomised studies and cannot prove that the improved outcomes observed are causally related. If it is permissible to claim 50% of the credit, then with some conservative assumptions about the benefits and allowing for both local and central costs, it is possible to do a 'back of the envelope' calculation. This would suggest a cost/QALY might be of the order of £50–£400. This would be an order of magnitude less than the costs associated with most new pharmaceuticals.

3 Finally, most scientific research studies produce results of which perhaps 10% or less will ever enter clinical practice and only then with a delay of many years. Well-conducted comparative audit studies will all yield results that can improve practice and will do so within months.

Conclusion

Comparative performance assessment is beginning to provide data on the way that clinicians handle their clinical work. It can already be used as part of clinical appraisal and is likely to develop into more sophisticated tools over time. Sharing such data with health service management should improve both effectiveness and efficiency, and sharing with patients/public will build confidence in the service.

It is a relatively cheap form of quality assurance that will become easier as IT develops. As a research exercise it can be justified as one that seeks not to achieve an impact factor but to have an impact.

References

1 www.bristol-inquiry.org.uk/final_report/
2 RCP and National Asthma Campaign. Guidelines for the management of asthma, British Thoracic Society, Kings Fund. *BMJ* 1990;**31**:653–7.
3 www.rcplondon.ac.uk/college/ceeu/ceeu_guidelinesdb.asp
4 Leng G, Pearson M, MacBeth F. *Formulating consensus guidelines in delivering quality in the NHS*. Oxford: Radcliffe Medical Press, 2004.
5 Thompson S, Peake M, MacBeth F, Pearson M (eds). *Lung cancer: a core data set*. London: RCP, 1999.
6 Gompertz P, Irwin P, Morris R, Lowe D *et al.* Reliability and validity of the Intercollegiate Stroke Audit Package. *J Eval Clin Pract* 2001;7:1–11.
7 Birkhead J, Pearson MG, Norris R, Rickards A, Georgiou A. National Audit of Acute Myocardial Infarction: A new development in the audit process. *J Clin Excellence* 2003;**4**:379–85.
8 ISIS-2 Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both or neither among 17,187 cases of suspected myocardial infarction: ISIS-2. *Lancet* 1988;**ii**:349–60.
9 www.icnarc.org/Audit.htm.
10 Birkhead JS, Walker L, Pearson M, Weston C *et al.* Improving care for patients with acute coronary syndromes: initial results from the National Audit of Myocardial Infarction Project (MINAP). *Heart* 2004;**90**:1004–9.