Evaluation of medical audit

Michael B Robinson

Abstract

Objective – To review current knowledge of the effectiveness of medical audit programmes as a whole and of specific interventions within these programmes, as a means of changing clinical behaviour.

Criteria for inclusion and exclusion of published reports – Articles listed on Medline from 1985-92 with key words “quality assurance” or “medical audit”, and “evaluation” and relevant references from these articles, and from recently published reviews and reports on medical audit, were included. Excluded were simple descriptions of audit activity, replication of previous work, and publication in a language other than English.

Results – Evaluation of entire programmes of medical audit is unusual. Most reports concern specific interventions and focus particularly on the scientific and technical aspects of quality. These interventions may be classified by the means through which they attempt to achieve desired changes: patient characteristics; physician characteristics; administrative and organisational structures; and financial incentives.

Conclusions – Knowledge about effective methods of bringing about specific changes in clinical behaviour is rudimentary. Impact is highly dependent on local factors, so generalisation of results to other settings is difficult. More qualitative research is needed to define the local factors which influence results.

(The Epidemiol Community Health 1994;48:435–440)

The meaning of evaluation

In common usage, evaluation means “to determine the amount, value or significance of”. A more technical definition is: “a process that attempts to determine as systematically and objectively as possible the relevance, effectiveness and impact of activities in the light of their objectives”.

Other definitions vary in detail, but all suggest the need for clear objectives for the activity under consideration.

In the case of medical audit, objectives are imprecise or controversial. “To improve the quality of care doctors provide for their patients” is simple, but of limited value in the absence of a common definition of quality. Guidance from the DH itself or the Royal Colleges has similar limitations. Attempts to define more precise objectives, such as improvements in health status or reductions in adjusted mortality rates, run counter to clinicians’ focus on individual patients.

In light of the lack of clear objectives, evaluation of entire programmes of medical audit is difficult. Most published reports are based on surveys of those responsible for implementation: quality assurance directors, medical administrators, and chairpersons of medical audit committees. Judgement of effectiveness is subjective and the respondents may have a vested interest in producing positive results. However, achievements seem limited: in the United States, with a long tradition of formal quality assurance, only 20% of respondents perceived changes in clinical practice attributable to audit.

Rather than attempting to evaluate audit at a “macro” level, most research has focussed on specific interventions which may be used in conjunction with audit to stimulate specific changes in clinical practice. Three types of evidence have encouraged these developments:

1. The existence of variations in practice which cannot be explained by patient characteristics, or simple resource constraints.
2. The failure of simple educational interventions to modify individual practice.
3. Changes in clinical behaviour observed when payment schedules or other regulations are introduced.

There are two key questions. (1) Which interventions are effective means of changing...
practice, and under what circumstances? (2) How does their cost effectiveness compare with investment in direct patient care?

Scope and limitations of published reports
Several previous reviews provide a general overview of strategies for changing clinical behaviour.26-28 Others consider specific approaches such as education29-31 or feedback,32 or particular fields of application such as cost containment of laboratory testing,33-34 or prescribing.35,36

All but one of these reviews was written in the US or Canada, and some of the methods described are not readily transferable to the UK. The aim of this review is to focus on methods which would be practical for use in conjunction with medical audit as currently organised in the NHS.

Such methods have conventionally been classified into four types: feedback; education; administrative; and financial,39 based upon early cost containment studies in the United States. Clinical behaviour as a whole is influenced by a wider range of factors, however, so a more general system has been proposed.32 This recognises four groups of factors, each of which has been shown to influence clinical behaviour and so can potentially be manipulated to achieve desired changes. These four groups are as follows:

(1) Patient characteristics, such as expectations of treatment and level of knowledge;
(2) Physician characteristics, such as seniority and educational achievements;
(3) Administrative and organisational structures, such as local policies and guidelines, peer review, and external monitoring mechanisms; and
(4) Financial incentives and disincentives, such as fee schedules and payment systems.

The most successful audits use an approach which modifies a combination of these factors, so the division into four groups is somewhat artificial. The scheme is useful, however, for describing the full range of possible interventions.

Limitations
Three general limitations of published reports need to be kept in mind. Firstly, the narrow scope. Almost all studies focus on the scientific and technical aspects of quality, rather than equity or humanity. A study in Minneapolis of 60 000 criteria used in 448 audits found that only 3% related to psychosocial aspects and 4% to communication with patients.37 The interpersonal aspect of health care, despite being important to patients, has rarely been investigated.38

A second problem is methodological weakness. Although publications are extensive, most reports are descriptive.39 Even when evaluation has been attempted, it generally consists of "before and after" observations. Studies in which the baseline data are collected retrospectively and the "re-audit" is conducted prospectively39,40 are difficult to interpret because part of any change is due to a simple Hawthorn effect.32 The most reliable evidence comes from randomised controlled trials, but these are uncommon. In a review of continuing medical education, only seven of 248 articles met the authors' reliability criteria,29 and in a similar review of computer assisted feedback only 14 of 135 were satisfactory.30 Restricting consideration to only those interventions which have been so tested may, however, exclude some which are effective, as not all are suitable for randomised trials.41

The third limitation concerns what is actually being measured as an indicator of effectiveness. Davis and Haynes recognised four levels: (1) participants' perception of change; (2) change in knowledge; (3) change in clinical process; and (4) change in patient outcome.42 Lomas showed how clinicians' perceptions of change were greater than actual changes in performance.43 Similarly, the link between change in knowledge and change in behaviour is often inconsistent. The final and most important step, to demonstrate changes in patient outcome, is the most difficult, and has been shown in only three of 13 randomised studies designed to measure this to date.27

With these limitations in mind, the range of possible interventions will now be considered, concentrating on evidence for effectiveness. Little is known about cost effectiveness.44 Most reports come from the United States, where rising costs have stimulated intense interest in such interventions.

Patient characteristics
The idea that medical practice is heavily influenced by the patients' own knowledge and expectations is widely acknowledged, especially in primary care.44 Attempts to manipulate this consciously to change clinical behaviour are, however, rare. An example showing that this method has considerable potential is reported from Switzerland, where a media campaign aimed at the general public in one locality was associated with a 25% decrease in the hysterectomy rate which was sustained over the next three years.45

Such approaches can only be used to implement change as part of a specific audit when patient behaviour is a significant determinant of clinical performance. An example is the possible use of a public campaign to increase the early use of thrombolysis for suspected myocardial infarction.46 Evaluation of such methods is difficult because of their diffuse nature27 but this should not exclude their consideration, especially as they may be relatively inexpensive.

Physician characteristics
Experience and educational attainment are associated with variations in clinical practice.47 Attempts to use these to modify behaviour can be divided into those which include detailed advice on the management of specific clinical
situations (guidelines), and those which adopt a less prescriptive approach.

NON-PRESCRIPTIVE EDUCATION
The strength of this approach is that any change in practice ought to be long lasting, in contrast to approaches based on feedback, where continual reinforcement is likely to be necessary.

Most reported evaluations are positive but methodologically unsound.43 The small number of carefully designed studies are mostly negative,29 but these have not always used the most effective teaching methods. Four basic approaches have been described: (1) sending material by post; (2) traditional didactic lectures and presentations; (3) individually tailored instruction using small groups; (4) one to one contacts.

In general, the first two methods are ineffective, the third has a small effect, and the fourth is the only consistently effective one. The observation that pharmaceutical marketing companies spend most of their resources in this way is suggested as empirical support.68 Surveys of the diffusion of new technologies among doctors,63-69 and polls asking what sources of information are perceived to change practice,70 also emphasise one to one learning from colleagues.

Two factors which may be important in determining effectiveness are the status of the educator as perceived by the participants, and the environment in which the education takes place. Contact with physicians is more likely to change prescribing than similar interventions by pharmacists.51 In Denmark, the same senior physician gave lectures on antibiotic prescribing in two localities, as a follow up to the dissemination of written guidelines.54 In the first locality, the lectures were organised by the local medical society and sponsored by the pharmaceutical industry. Subsequent antibiotic use was no different to that in localities which received written guidelines only. In the second locality, lectures were held in the local microbiology department and significant changes in prescribing were observed. Physical location may be an important determinant of effectiveness.

Combining education with feedback is frequently used to try to improve effectiveness.59-57 In such cases, persistence of the change in practice after feedback has been stopped would suggest that the educational component had been effective. This is unusual:26-59 only one to one approaches show prolonged effects.60 In summary, education is not as effective as intuition would suggest. It must be used in combination with other strategies or delivered in a relatively expensive way by one to one contact.

GUIDELINES
The production of guidelines has become a major industry – a recent US survey identified 26 medical organisations which are currently active in this field, and a further 10 with plans for the future.61 Some programmes involve multiple stages, and resulting guidelines may be used by third party payers, such as health insurance companies, to assess claims.62 Official guidelines, however, may not be widely adopted.63

Achieving consensus over the detail of guidelines is an important step.64 Consensus development conferences are commonly used, but their effectiveness in changing behaviour is uncertain.6566 Factors which are associated with effective guidelines are application to radiology or pathology services6768 and an emphasis on multidisciplinary involvement in design and implementation.6970 A randomised trial in Canada showed that education by locally chosen, respected peers was much more effective than a formal audit programme as a means of implementing national guidelines.71

US authors disagree about whether external guidelines, without local modification, can provide an effective means of changing practice.6572 Consideration of the entire “life cycle” of new fashions in clinical practice, from early innovation to obsolescence, suggests that formal expression as “guidelines” tends to come too late to be useful.73 The considerable resources consumed in guideline production may not be justified.

ADMINISTRATIVE AND ORGANISATIONAL FACTORS
Modification of administrative or organisational systems is an important alternative to education. The simplest way to modify behaviour is to constrain it by enforcing local policies and procedures. Simple devices such as the posting of guidelines on notice boards and the introduction of local formularies have been shown to be effective.1455 The scope for such prescriptive tactics, however, is limited. To justify these measures, feedback of information on performance is a central part of most organisational approaches. There are four courses of action:

(1) Peer review of individual cases using implicit criteria;
(2) Explicit audit (predefined criteria);
(3) Expert assistance (computers or people); and
(4) External systems (covering several hospitals).

PEER REVIEW
Peer review is used here to mean retrospective review of individual episodes of care without explicit criteria. The lack of explicit criteria means that the entire review process must be conducted by clinicians themselves, and so either the number of cases examined is relatively small or judgements are made from abstracts prepared by clerical staff.76

Although this approach has been discredited in the US for being too subjective,7778 it has been widely adopted in the UK as the easiest way of conforming with the government’s demand that all clinicians take some part in regular audit. No randomised trials have been conducted to assess effectiveness, but “before and after” studies show that the quality of note
taking improves, mostly in the first year after introduction.\textsuperscript{79,80} Despite better recording, no improvement in the process or outcome of care could be shown for the one particular condition studied (asthma).\textsuperscript{81}

The limited effectiveness of peer review may be related to the fact that it is perceived as repetitive and boring.\textsuperscript{80,82} A focus on the completion of records might even lower the quality of care, by diverting time and energy from direct patient care.\textsuperscript{83,84}

**EXPLICIT AUDIT**

The use of explicit criteria for retrospective audit is a tried and tested method in the US, having been included as part of the programme of Peer Standards Review Organisations (PSROs).\textsuperscript{85} Evaluations have produced mixed and confusing results.\textsuperscript{27}

One reason for this is that explicit audits have been used for two separate purposes: firstly, to discourage overutilisation, and hence limit costs, and secondly to expose and correct underutilisation and hence "improve quality". Programmes with the second type of objective consistently report positive effects, from a variety of situations – rural practices,\textsuperscript{86} urban ambulatory care,\textsuperscript{87} and accident and emergency.\textsuperscript{88}

A drawback of most of these reports is that measurement of effectiveness is based on what is recorded in the case notes, but in at least one instance,\textsuperscript{87} a large part of the observed improvement was accounted for simply by more complete recording. A recent literature search\textsuperscript{27} identified only three randomised trials that had measured outcomes; of these only one showed improvement, which was minor.\textsuperscript{88}

Behavioral theory suggests that feedback will be most effective if given as soon as possible after performance, and this has become known as "concurrent" feedback. A recent review of five articles using such methods all reported positive results.\textsuperscript{89} An earlier randomised controlled trial, however, was negative.\textsuperscript{89}

Use of computerised medical records and explicit algorithms allows "concurrent" feedback to become truly simultaneous with clinical performance by generating reminders for individual consultations. McDonald demonstrated that the effectiveness of this method was not due to an educational effect because original behaviour returned when reminders were withdrawn.\textsuperscript{89,90} The marginal costs of generating reminders if a computerised record system is already in place are low, and in a rare example of economic evaluation, a satisfactory cost-benefit relationship has been demonstrated for the treatment of streptococcal sore throats.\textsuperscript{91}

Another method of improving the effectiveness of explicit audit is to actively involve participants in the choice of criteria.\textsuperscript{92,93} Individual participation, directly rather than through members of one's peer group, seems important. Anderson showed that participating GPs changed their use of digoxin, whereas their partners did not.\textsuperscript{94} The North of England General practice study showed that groups of trainers modified their behaviour to achieve self determined standards, but not those worked out by another group.\textsuperscript{95}

In summary, the main factors associated with effective explicit audit are:

- **(1) Aiming at correction of underutilisation rather than overutilisation;**
- **(2) Rapid feedback or individual patient reminders;**
- **(3) Involvement of participants in audit design.**

**EXPERT ASSISTANCE**

Rather than provide reminders or feedback, another approach is to provide expert support either in the form of specialist staff or from a computer. Non-medical staff have been successfully employed to query investigation requests and prescriptions,\textsuperscript{96} but this might be regarded in the UK as an unacceptable infringement of professional boundaries by unqualified staff. More refined versions have used clinical pharmacists to suggest modifications to prescriptions\textsuperscript{97,98} or physiotherapists to query lung function tests.\textsuperscript{99}

A less confrontational approach is to use a computer rather than a person to deliver "expert" help, for example to assist in the diagnosis of acute abdominal pain. Initial results in the UK showed a 50% reduction in the negative laparotomy rate\textsuperscript{100} and were reproduced by a multicentre study.\textsuperscript{101} The method has not, however, been widely implemented. Randomised trials of expert systems for adjusting dosages of warfarin and lignocaine in the US also showed positive results.\textsuperscript{27} In general, this strategy seems underexploited, but positive evaluations do not necessarily mean that clinicians would use expert systems if they were installed.\textsuperscript{27}

**EXTERNAL SYSTEMS**

In the US, the most successful external audits have been undertaken by medical societies, particularly with regard to elective surgery.\textsuperscript{102,103} Simple feedback of rates per thousand population was associated with a narrowing of the variation between hospitals due to a reduction in the higher rates, but catchment populations were unusually well defined and successful replication elsewhere has not been reported.

The desire to contain costs in US health care has led to the development of external regulation specifically to control overutilisation, a process known as "utilisation review."\textsuperscript{104} Examples of these methods are "pre-admission certification", whereby a claim for hospital care will not be paid unless prior authorisation for the episode has been obtained from the third party, and mandatory second surgical opinions. UK purchasers could feasibly introduce equivalent measures, particularly for extracontractual referrals.

Recent evaluations show appreciable cost savings associated with utilisation review,\textsuperscript{105,106} but these studies are not designed to measure any deterioration in patient outcomes associated with successful cost containment. Another limitation is that the findings refer to
FINANCIAL INCENTIVES

The potential ability of economic incentives to modify clinical behaviour has been demonstrated by several "natural experiments" where changes in physician payments were followed by changes in clinical behaviour.20 Deliberate manipulation of such systems, however, to produce a specific change in behaviour is "an entirely unresearched area."21 The only controlled trial of direct financial incentives controlling the costs of investigations by junior hospital doctors showed no difference between control and intervention groups.107 A survey of medical records officers in the US to determine what methods were used to encourage clinicians to complete case notes on time, however, found that financial incentives were commonly used.108

At a national level, changes in payment systems such as the introduction of prospective payment have had a clear effect on clinical behaviour.109 In Denmark, the introduction of item of service payments to GPs was associated with changes in behaviour.110 The use of targets in the 1990 contract for GPs in the UK has had comparable effects.111 The release of hospital trusts from compulsory Whitley Council employment conditions may provide similar opportunities within hospitals.

Conclusions about effectiveness and cost effectiveness of methods for changing behaviour

Knowledge about effective methods for inducing specific changes in clinical behaviour is rudimentary.112-113 Impact is highly dependent on local factors, which are difficult to measure and adjust.

When effectiveness is uncertain, cost effectiveness, by definition, must also be so. Some individual studies contain details of costs and benefits, but the extent to which these findings can be generalised is doubtful. The scope for using economic methods of evaluation to justify investment in audit seems limited.

The studies reviewed here, however, suggest a number of interventions that may be effective under certain circumstances. More qualitative research is needed to define the local factors which determine effectiveness. At present, the statement that "realising the benefit of audit requires an act of faith"16 remains true.

44 Lomas J, Anderson GM, Dominick–Pierre K, Chambers...
52 Geertsma RW, Whitbourne SK, Parker RC. How physicians view the process of change in their practice behaviour. Journal of Medical Education 1982;75:72-52.
93 Palmer RH, Lewis TA. A randomised controlled trial of the sixteenth annual computer assisted audit program. Med Care 1985;23:751-70.
96 Finn AF, Valentien PN, Burke MD. Alteration of physicians' orders not physicians. JAMA 1988;259:2549-52.