Audit: the third clinical science?

Ian T Russell, Brenda J Wilson

What is science?
Science may be defined as the "pursuit of knowledge;" and the scientific method as "the principles laid down in science for performing observations and testing the soundness of conclusions." Different accounts of what these scientific principles are, or should be, have been popular at different times. At the birth of scientific inquiry the dominant philosophy was rationalism built upon a framework of deductive logic. The doctrine of empiricism then emerged with inductive logic forming the basis of reasoning. The inability of inductive reasoning to prove cause and effect led to the development of probabilistic logic. This in turn gave way to the philosophy of Karl Popper that knowledge accumulates only by falsification. Thus scientific principles are not immutable but reflect the current scientific consensus.

Popper's philosophy is based on testing hypotheses. According to Popper, hypotheses are never proved but are supported to the extent that they survive attempts to falsify them by confronting them with empirical reality. To many this is the only true way of conducting science. Although this philosophy dominates current scientific thinking, it is not without its critics. Although opinion about the nature of true science is therefore not unanimous, there is broad agreement on the basic principles of the scientific method: firstly, hypotheses should be grounded in empirical data; secondly, hypotheses should be tested by other empirical data; and thirdly, the scientist must record how the testing was done. Almost all clinical research (that is, research in medicine and health care) follows these three scientific principles, usually in five practical stages: 0 Summarise existing knowledge in the form of a testable hypothesis 1 Design unbiased and precise methods for testing this hypothesis 2 Collect valid and reliable data 3 Analyse these data carefully 4 Present clear conclusions (and submit them to peer review).

Two types of clinical science?
There are at least two distinct types of clinical research: explanatory and pragmatic research. Explanatory research is designed to generate knowledge that explains biomedical phenomena - for example, the aetiology of a disease or the pharmacology of a drug. When explanatory research is conducted through randomised trials these "explanatory trials" seek to reproduce the strictly controlled conditions of the typical laboratory experiment: treatments are rigidly defined, eligibility for both treatments (and thus the trial) is specified a priori; patients who withdraw from a treatment are withdrawn from the analysis; and the emphasis of that analysis is on statistical testing of a hypothesis.

In contrast, pragmatic research is designed to generate knowledge that improves resource allocation decisions - for example, whether to initiate a screening programme or to recommend a drug. When pragmatic research is conducted through randomised trials these "pragmatic trials" seek to reproduce the normal clinical circumstances in which the ensuing decisions will be implemented: "treatments" are flexibly defined, so that they can be adapted to individual patients; eligibility for both treatments (and thus the trial) depends to some extent on clinical judgement; patients who withdraw from a treatment remain in the trial for analysis (which is therefore described as analysis by intention to treat); and the emphasis of that analysis is predominantly on statistical estimation - for example, by using confidence intervals.

To illustrate the difference between explanatory and pragmatic research we take the example of prophylactic β blockade after myocardial infarction, which has been investigated in many studies over the past 20 years. Exploratory research in the laboratory has shown that β blockers protect against the adverse effects of myocardial ischaemia. Pragmatic research, in the form of a large (but simple) randomised trial, has shown that these effects translate into substantial clinical benefits.

Healy has argued that while explanatory research fulfils the strict Popperian criterion for science, pragmatic research does not and is better regarded as technology. However, pragmatic research does follow the three scientific principles described above. Though accepting the distinction that Healy was seeking to draw, we judge that it is more consistent with the current consensus among health professionals to describe explanatory research as biomedical science (that is, pure clinical science) and pragmatic research as health care science (that is, applied clinical science).

What is clinical audit?
The government white paper Working for Patients described audit as "systematic, critical analysis of the quality of medical care, including the procedures used for diagnosis and treatment, the use of resources and the resulting outcome for the patient." To provide a working definition, this useful
description needs to be augmented by four of the key elements of audit proposed by the Standing Committee on Postgraduate Medical Education:

A Audit should be directed at quality of care
B Audit should include the setting of standards
C Audit should compare performances with these standards
D Audit should lead to beneficial change.

The first of these elements represents the purpose of audit; the last three represent the three essential steps of the audit cycle – namely, set the standard; observe practice and compare with (this) standard; and implement change. Where does clinical audit, thus defined and elaborated, fit into our taxonomy of biomedical and health care science?

Rather than argue from first principles, we begin with another case study of prophylaxis after myocardial infarction. Eccles and Bradshaw, concerned about the quality of coronary care in the north of England, surveyed a 50% random sample of general practitioners in the region to ask whether the subsequent patient discharged from hospital after myocardial infarction was taking prophylaxis. They had derived a simple standard from published reports: that all such patients should be taking both a β blocker (unless they had diabetes mellitus, peripheral vascular disease, or obstructive airways disease) and aspirin (unless they were sensitive to aspirin, or had a duodenal ulcer, or were taking an anticoagulant). They measured clinical performance by recording that the treatment of only 109 (41%) of the 267 patients reported to them met this standard. Finally, they sought to stimulate beneficial change, notably by publishing their findings in the BMJ, the scientific medical journal with the widest circulation in the United Kingdom.

Because the hospital physicians whose performance was measured were not party to the study it was not described as audit. It was reported in the general practice section of the BMJ rather than the audit section. Nevertheless, it does contain the four key elements of audit listed above. That beneficial change was sought through publication, one of the least effective strategies for closing the gap between performance and standard, does not undermine our view that this study provides a good model for audit. Yet the study also follows the three basic scientific principles. Firstly, the authors derived a hypothesis – that secondary prophylactic drugs were not being widely used – from an empirical preliminary study in Tyneside. Secondly, as we have shown, they tested this hypothesis against new empirical data, using the five practical stages described above. Thirdly, they recorded how this testing was done with enough rigour and detail to satisfy the referees of the BMJ.

Most reputable audits also adopt the three basic scientific principles. We argue that clinical audit therefore fulfils the criteria for consideration as a branch of clinical science. There is, however, a far more cogent reason for treating audit as science and demanding that it fulfil all the desiderata that we have proposed for a scientific activity – namely, an appropriate hypothesis, unbiased sampling, valid and reliable data, careful analysis, and clear presentation. Unless audit is seen by the health professions as a rigorous discipline that itself sets high standards in the pursuit of worthwhile benefits for patients and professionals it will surely fail to deliver those improvements in quality of which published reports show it is capable.

That said, we acknowledge that the science of clinical audit is logically distinct from both biomedical science and health care science. Audit is designed to generate knowledge that enhances the quality of health care (including the quality of individual clinical decisions) rather than improving resource allocation decisions or even explaining biomedical phenomena. Of the many different ways in which this distinction is manifest, one is paramount. Biomedical research aims at drawing scientific inferences about the nature of biological phenomena and health care research about the effectiveness of clinical activities or policies; but both seek to draw their inferences over a representative selection of patients or doctors. In contrast, audit seeks to draw scientific inferences about the performance of individual doctors or groups of doctors; it seeks (eventually) to cover a representative selection of their clinical activities. We judge that clinical audit is most likely to succeed if it is recognised as the third clinical science – with its own theories, techniques, and literature.

Audit as rigorous science

We have described five practical stages of clinical science and three essential steps of the audit cycle. If audit is indeed to be recognised as the third clinical science it is timely to integrate these two components into a coherent scientific basis for audit. In doing so, we also take account of a recent review of the effectiveness of “guidelines,” a common synonym for standards. By restricting attention to evaluations of clinical standards based on rigorous statistical designs, Russell and Grimshaw showed that effectiveness depends not only on how standards are developed but also on how they are disseminated and implemented. This leads us to propose a cycle of nine distinct steps for conducting scientific audit:

A Choose a general topic for audit, and a specific hypothesis to be tested.
B1 Develop a standard.
B2 Disseminate the standard.
B3 Implement the standard.
C1 Design unbiased and precise methods for sampling patients.
C2 Collect valid and reliable data on performance.
C3 Compare these performance data with the standard by careful statistical analysis.
C4 Feed a clear summary of this comparison back to participants.

D Ensure that this process generates beneficial change.

As steps C1, C2, and C3 are directly analogous to the corresponding stages of clinical research the established methods of ensuring scientific rigour can be adopted with little change. As the remaining six steps are much more specific to audit they represent the main topics on which research into audit will need to focus over the next few years. Fortunately, literature reviews have already begun to identify effective audit techniques and to propose an agenda for future audit research. Furthermore, the growing British portfolio of audit textbooks relating to audit both in hospital and in general practice provide sound practical advice, not least on the need for rigour in one form or another.

In these circumstances, we conclude by using another case study to illustrate how rigour is needed at each of the nine steps of a scientific audit project. We return to the topic of prophylaxis after myocardial infarction. Although the study reported by Eccles and Bradshaw was admirable in many respects, it was not designed as audit and sought beneficial change only through publication. We therefore ask how audit in this field could be made effective. In particular, we explore how prophylaxis after myocardial infarction might be audited within a single district general hospital.

A CHOOSE A GENERAL TOPIC AND A SPECIFIC HYPOTHESIS

The combination of unequivocal evidence from two of the largest randomised trials ever conducted and strong evidence that prophylaxis after myocardial infarction is not widely practised makes this an especially appropriate and potentially productive topic for audit. Suppose that the physicians conducting this audit adopted a standard slightly less strict than that proposed by Eccles and Bradshaw: that 90% of patients discharged with confirmed or suspected myocardial infarction should have been prescribed both a β blocker (in the absence of specified contraindications, including heart failure) and aspirin (in the absence of other specified contraindications, including heart failure). Then the proposed audit should test the hypothesis that the treatment of 90% of such patients meets this criterion.

B1 DEVELOP A STANDARD

The existence of unequivocal scientific evidence makes developing a standard much easier than usual. Nevertheless we suggest that the proposed standard should not be adopted until it has been thoroughly reviewed – for example, at a meeting of senior medical staff. There is published evidence that such a review increases the level of commitment to the standard.

B2 DISSEMINATE THE STANDARD

Published evidence also suggests that once formally adopted by senior staff, the standard should be disseminated to junior medical staff through an educational programme designed for this purpose – for example, seminars that present the evidence for prophylaxis after myocardial infarction and discuss how the standard should be implemented.

B3 IMPLEMENT THE STANDARD

Adherence to standards is much enhanced when doctors are alerted by patient specific reminders at the time of a consultation. This could be achieved by using coloured flags to identify the medical records of all patients with suspected myocardial infarction soon after admission. To make adherence as easy as possible each flag should ask four simple questions (and provide four boxes for the answers):

(1) Has this patient got any of the following contraindications to a β blocker?
(2) If no to question 1, has a β blocker been prescribed yet?
(3) Has this patient got any of the following contraindications to aspirin?
(4) If no to question 3, has aspirin been prescribed yet?

C1 DESIGN UNBIASED AND PRECISE METHODS

FOR SAMPLING PATIENTS

We judge that to ask the doctors whose performance is to be audited to identify relevant patients carries too great a risk of bias, either conscious bias or more likely, subconscious bias. To eliminate this risk we suggest that patients should be identified by independent professionals – for example, audit facilitators or ward secretaries; ideally a sample of identified patients should be validated by other independent professionals – for example, electrocardiography technicians.

If the hospital where the audit is being conducted serves a population of 250 000, an annual incidence of suspected myocardial infarction of two per 1000 and a hospital survival rate of 80% would yield 400 patients a year for audit. If the audit is to be carried out one year after the standard has been disseminated and implemented we need to decide what proportion of the corresponding 400 records should then be sampled at random for review. We should ask the senior medical staff how far below their standard of 90% adherence to the agreed drug treatment would practice need to fall before the difference became clinically significant. Suppose they consider adherence as low as 80% to be clinically important and that they would like to detect such a departure from their standard on at least 95% of the occasions on which it occurs. Then statistical calculations along the lines described by Moser and Kalton show that 133 records should be sampled.

Thus the audit should be based on a one in three sample of the 400 eligible patients. To encourage continuous compliance with the standard this sample should not be drawn until the end of the year. At that time every
third patient should be sampled from a list arranged in order of admission; to avoid bias the first patient should be sampled using a random number.21

C 2 COLLECT VALID AND RELIABLE DATA ON PERFORMANCE

The questions on the coloured flags used to identify the medical records of relevant patients would also enable an audit facilitator to see at a glance whether drug treatment had met the agreed criterion; he or she would need only to check that question 1 or question 2 (of subsection B3) had been answered "Yes," together with question 3 or question 4. To ensure that these answers were reliable it would be prudent to corroborate them against the medical record to which they were attached. To ensure that they were valid, in the sense that the risk of a second myocardial infarction was reduced by good patient compliance with the drugs prescribed, would be more difficult. We suggest that a brief postal questionnaire to the patient's general practitioner three months after discharge would be a cheap and sufficiently effective means of achieving this; the questionnaire would ask the same four questions (with the list of contraindications to β blockers extended to include side effects).

We are committed to the principle that whenever feasible and necessary audit should monitor patient outcome as well as medical process.32 This audit is unusual in that the measurement of definitive patient outcome is neither feasible nor necessary. To detect a reduction in patient mortality would require far more patients than the maximum of 400 available. Fortunately, the unequivocal evidence about the effectiveness of prophylaxis19 31 means that good patient compliance with the drugs prescribed is a good "intermediate outcome," in the sense that it is a good predictor of definitive outcome. In less favourable circumstances, when the medical process advocated in the adopted standard had not been validated scientifically, a scientific audit would need to augment measures of medical process with valid and reliable measures of patient outcome.33 34

C 3 COMPARE PERFORMANCE DATA WITH THE STANDARD

In this audit the statistical analysis is unusually simple: if fewer than 114(86%) of the desired sample of 133 patients are prescribed drugs consistent with the agreed criterion then the performance of the physicians being audited is statistically (and clinically) worse than the standard of 90%.19 21 35

C 4 FEED A CLEAR SUMMARY OF COMPARISON BACK TO PARTICIPANTS

Although the evidence on the effectiveness of different methods of presentation is flimsy,18 our own experience suggests that attractive graphical summaries are more effective in stimulating beneficial change than routine numerical summaries.36 Such graphs should compare the actual performance of all physicians within the hospital with the agreed standard – for example, by showing the estimated percentage adherence, augmented by a confidence interval to show the variability of this estimate.39

D ENSURE THAT THE PROCESS GENERATES BENEFICIAL CHANGE

Even an audit as rigorous as we have proposed might not bring actual performance closer to the agreed standard without careful attention to the techniques of managing change.33 The importance of these techniques within a rapidly changing health service is increasingly recognised, especially in general practice.28 38 39

Conclusion

The hypothetical audit that we have described is simple, cheap, and potentially very effective. Because it pursues scientific principles and methodological rigour it avoids an unnecessarily large sample and unnecessary complexity in the form of a computerised database. Because it adopts both a proved standard10 31 and audit methods whose effectiveness have been established by sound evaluation17 it could greatly improve the quality of coronary care.

In summary, we believe that if clinical audit is accepted and prosecuted as the third clinical science, it has the potential to deliver substantial benefits to patients and health professionals.

We thank Drs Martin Eccles, Colin Bradshaw, and John Rawles and colleagues in Aberdeen for helpful comments on an earlier draft of this paper. The Health Services Research Unit is funded by the chief scientist office, Home and Health Department, Scottish Office. However, the Scottish Office is not responsible for the opinions expressed in this paper.


Audit: the third clinical science?

17 Russell IT, Grimshaw JM. The effectiveness of referral guidelines: a review of the methods and findings of published evaluations. Aberdeen: Health Services Research Unit, 1991. (Occasional paper 8.)