Detecting the effects of thromboprophylaxis: the case of the rogue reviews

Mark Petticrew, Susan C Kennedy

"Watson, Dr Watson!" I looked up at the muffled sound of Sherlock Holmes's voice as he stood at the window, gesticulating with his Stradivarius. Two hours of relentless arpeggios had finally ceased, and I gratefully removed the plugs of cotton wool from my ears. "A distinguished visitor is about to request admittance," he observed. "Would you kindly ask Mrs Cochrane to show him up?"

A few moments later a tall, bewhiskered gentleman, with an enormous portmanteau and a general air of exasperation, entered Holmes's study. He introduced himself as Professor Legge, an orthopaedic surgeon.

"Mr Holmes, only you can end this madness!" he moaned, sinking into the nearest armchair. Holmes's hawk-like eye ranged over his visitor, and I knew that the great detective was about to presage the discussion with a display of his deductive skills. "Well, Professor Legge," he began authoritatively, "I trust your search for systematic reviews on Medline this afternoon was productive?" Legge looked startled.

"Good Lord, Mr Holmes, how could you possibly know that I have just spent hours . . . searching for . . . ?" He began to swab his face with an extravagantly large handkerchief.

"Simple, my dear Legge. The light coating of dust on your face would indicate that you have attracted an electrical charge, caused, I suspect, through hours of vigilant study at a computer monitor. That, and your rather glassy stare." Legge nodded silently, while I enquired: "But how on earth did you deduce the reason for his search?"

"Are you unaware that the use of meta-analyses in the pursuit of effective health care is well established in your profession, Dr Watson?" he retorted. I was abashed. Holmes continued: "Well, I too am an exponent of evidence based methods. Naturally I am aware of their value in medical practice." Professor Legge grunted sceptically and began rummaging about in his bulging portmanteau.

"Well, Mr Holmes," he said, thumping a large volume of paper on to the occasional table. "I'd think these blasted meta-analyses were valuable too if they agreed among themselves." He leaned forward, almost triumphant. "But, you see, they don't. I've been trying to get some straight answers on effective surgical thromboprophylaxis all afternoon, and frankly, given the conflicting information in this lot," he jabbed his finger at the innocent pile of paper, "I'd be forgiven for mistrusting everything except what I've seen with my own eyes."

Summary points

Holmes and Watson are visited by an exasperated Professor Legge

Thumping a large pile of meta-analyses on the table, Legge explains that he's been trying to get straight answers on effective surgical thromboprophylaxis—without success

Over tea Holmes dissects Legge's problem: the principal difficulty, he asserts, lies in too much evidence. The vital evidence is obscured by erroneous information—caused on this occasion by flawed methods

After contemplating the evidence Holmes concludes that mechanical methods are the answer to preventing deep vein thrombosis and that a comprehensive literature search, explicit inclusion criteria, detailed assessment of quality of studies, and appropriate methods of pooling the data are the key to a good review

"Ah, yes," Holmes reflected, "thromboprophylaxis in general and orthopaedic surgery is a rather vexed issue at present, is it not?"

I was not surprised to hear Holmes speak knowledgeably on thromboprophylaxis. I knew that the months following his apparent death at the hands of the evil Professor Moriarty were spent in chemical research in a French laboratory. I suspected now that this included work on pharmacological thromboprophylaxis. Moreover, his medical interests were wide enough for him to be claimed, in later years, by neurologists,[1] anaesthetists,[2] dermatologists,[3] and ophthalmologists[4] as one of their own. I did not doubt that those interests extended to surgery, though I confess I was surprised at his grasp of information technology, which had not, as yet, been invented.

"Perhaps, you would oblige us with the background to the case Professor Legge?" Holmes suggested. The professor nodded and leaned back into his armchair.

Professor Legge's problem

"As you may know, patients undergoing major surgery are at risk of thromboembolic disease. Half of orthopaedic patients receiving no prophylaxis develop deep vein thrombosis,[5] and almost a quarter of deaths after orthopaedic surgery have been attributed to pulmonary embolism.[6] "

NHS Centre for Reviews and Dissemination, University of York, York YO1 5DD

Mark Petticrew, *research fellow*

12 Park Crescent, York YO3 7NU

Susan C Kennedy, *adult basic skills tutor*

Correspondence to: Dr Petticrew mp25@york.ac.uk

BMJ 1997;315:665–8

"But surely most of these cases of deep vein thrombosis are clinically unimportant?" I interrupted.

"It is true that most cases are subclinical and resolve spontaneously, Dr Watson, but there is still significant morbidity associated with the condition," Legge rejoined. "Besides, cold hearted as it sounds, the treatment of problems associated with deep vein thrombosis costs almost half a billion guineas annually.[7] Imagine, moreover, the patient who is harmed, sometimes fatally, by thrombosis as a result of an operation, the purpose of which is to cure."

"Remind us, Legge, how the problem can be prevented," Holmes requested, growing more intrigued.

"Broadly, we have either the pharmaceutical methods (such as heparin, low molecular weight heparin, aspirin, and warfarin) or the mechanical methods (such as elasticated stockings). Prevention seems to represent the best management strategy, and yet effective thromboprophylaxis does not appear to be used routinely in high risk surgery."[6]

Holmes regarded Legge thoughtfully. "I take it that this is not simply due to some oversight on the part of British surgeons?"

"Certainly not! There are two principal reasons. Some surgeons believe that while prophylaxis is effective against deep vein thrombosis, its benefit in preventing pulmonary embolism has not been proved."[8]

"And the other reason?"

"Safety, Mr Holmes. Surgeons are concerned about the risks of major bleeding and haematoma associated with pharmacological prophylaxis. If you've had a patient develop a major bleed during an operation, it's not something you forget. Now that's evidence." Holmes looked up sharply.

"But systematic reviews were developed to resolve just this kind of uncertainty. Are there many in this field?"

"Nearly two dozen,"[5 9-30] answered Legge ruefully. "And this, Mr Holmes, is where my faith in meta-analysis wavers. Disparity at all points of the compass! I'll give you an example: this one shows that in total hip replacement dextran, heparin, low molecular weight heparin, elasticated compression stockings, and warfarin all reduce the incidence of deep vein thrombosis, while aspirin, heparin, low molecular weight heparin, elasticated compression stockings, and warfarin all prevent pulmonary embolism.[21] And here's another showing that low molecular weight heparin is best."[20]

"Well, that's clear enough," I suggested blithely, "all methods work better than nothing at all." Legge gave me a rather withering look before returning to his notes.

"And here's one examining fatal pulmonary embolism after hip replacement, showing that none of these methods works better than no prophylaxis.[29] And there's another reason for my mistrust of meta-analyses." He leaned forward conspiratorially and continued sotto voce. "Some of my colleagues detect a sinister power behind many of the trials and reviews." Holmes sat bolt upright, his tufted eyebrows knitted together.

"Moriarty!" he breathed in a chilling voice. Legge looked at Holmes in surprise.

"Er . . . no. I was actually referring to the drug companies. Take low molecular weight heparin: some of my colleagues suspect that certain trials advocating its use were financed by the drug companies, in a dastardly attempt to make us use expensive drugs."

MARY EVANS PICTURE LIBRARY

"This is likely to be a three-pipe problem"

Holmes became thoughtful. "Well, are the reviews any clearer with regard to mechanical methods?" Legge again referred to his notes. "Well, Holmes, stockings do seem to prevent deep vein thrombosis in general surgery[31] and total hip replacement.[20] They may also prevent pulmonary embolism, but most trials have been too small to be sure. Many surgeons already use elasticated stockings, but they're not popular with patients—especially the chaps."

Legge by now was pacing anxiously. "You know, Holmes, I believe these meta-analyses are simply dredging the depths of scientific inaccuracy. Where will it all end, I ask myself? Meta-meta-analyses? Blast 'em all." Just then he noticed the formidable Mrs Cochrane, who had entered the library with a silver platter and was regarding him severely. "I'm sorry, ma'am, for my intemperate language," he mumbled.

"It's not the tone of your repudiation that worries me, Professor Legge," snapped Mrs Cochrane as she rattled the tea tray, "but its generalisability. 'Ere's your tea."

"Professor Legge," I ventured, seeing the heat of his passion pass, "do clinical guidelines provide any illumination on this most trying matter?" Legge's reply was weary. "Guidelines! Don't start me on the blessed guidelines! One lot cites as evidence a review using indirect comparisons.[32] Another is based on a mixture of conventional reviews, systematic reviews, and odd trials.[33] Another lot report that the incidence of fatal pulmonary embolism in high risk patients is about 1%,[34] though I've seen evidence to the contrary.[29] And as for the European Consensus Statement[35] ... well, Skrabanek had the right idea about consensus panels—'Nonsensus Consensus!' "[36]

I attempted to offer some bluff words of comfort. "Still, Legge, there's one consolation: no matter what you're doing, there's a meta-analysis to support it. Your practice is 100% evidence based!" A deathly silence pervaded the room. Holmes, Legge, and Mrs Cochrane stared coldly at me.

Holmes rose from his armchair and withdrew his pipe from the pocket of his dressing gown. "I shall now consider the case." He extracted an ounce of shag from his worn carpet slipper. "This is likely to be a three-pipe problem." He left the room with Legge's papers, and an hour passed before he returned.

Holmes's critical appraisal

Holmes sighed as he stood before us.

"I recall that a similar problem manifested itself in the case of the *Naval Treaty*,"[37] he began. "I suggested then that the principal difficulty lay in there being too much evidence: the vital evidence was obscured by erroneous information. Likewise with these meta-analyses."

"But why might their conclusions differ?" I pressed him.

"Consider their methodologies, Watson. One analysis might have included a different set of trials because they had employed different inclusion criteria, or simply a different search strategy." He picked up a handful of the professor's papers. "Furthermore, a closer examination shows that inappropriate methods of pooling data are sometimes used. For example, these reviews"[14 21 29 32]—he shuffled the offending papers—"have pooled data from similar treatment arms of trials and thereby compared the incidence of deep vein thrombosis under different forms of prophylaxis. The problem with that approach is that there may be differences between the trial populations. A similarly flawed method involves comparing the arms of different trials—say, the warfarin arm of one trial and the stockings arm of another trial—and using this as evidence that one method is more effective than another. Again, direct comparisons never took place, and the trials may differ in many respects."

"Such as the patients' characteristics, any other interventions which they may have received, and even the quality of the study," suggested Legge.

"Quite so, Professor Legge," agreed Holmes. "My original proposition was that systematic reviews reduce uncertainty. In this case they have increased it."

"Perhaps, then Holmes, you would venture a summary?" I suggested.

Holmes's conclusions

Holmes nodded. "Let us first confine ourselves to the essential facts. Surgeons are concerned about safety and effectiveness, and this has led to variations in practice as they seek to adapt conflicting evidence to the circumstances of particular patients. Some surgeons even believe that prophylaxis is not worth while, and that pulmonary embolism is rarer than often suggested."

"So what do you say to surgeons like myself who wish to use safe and effective prophylaxis?" pressed Legge.

"Simply this: mechanical methods. They are safe, and they have been shown to be effective in preventing deep vein thrombosis in patients at moderate and high risk.[20 30] They may even prevent pulmonary embolism. In short, gentlemen, a judicious use of mechanical methods and a suspicious mind regarding meta-analysis are the key to this whole affair. You've heard, of course, of the case of the Misleading Meta-analysis.[38] That, at least, should teach us caution."

"Do you imply, Holmes, that we should never trust another systematic review?" I interrupted in surprise.

"Far from it, Watson. Despite the recent plethora of antagonistic correspondence in learned journals, a methodologically sound systematic review remains the gold standard for the assessment of effectiveness." He gestured to the pile of papers on the table. "On the basis of what I have read here, there are four main indicators of a sound review: firstly, a comprehensive literature search; secondly, explicit, detailed, inclusion and exclusion criteria; thirdly, a detailed assessment of the quality of the included studies; and, fourthly, appropriate methods of pooling the data. The 'Sign of Four,' if you like, gentlemen!" He turned to me. "Is that succinct enough for your memoirs, Watson?" I nodded. "In fact it's . . . er . . . elementary!"

1 Cherington M. Sherlock Holmes: neurologist? *Neurology* 1987;37:824-5.
2 Maltby JR. Sherlock Holmes and anaesthesia. *Can J Anaesth* 1988; 35(1):58-62.
3 Dirckx JH. Medicine and literature: Sherlock Holmes and the art of dermatologic diagnosis. *J Dermatol Surg Oncol* 1979;5:191-6.
4 Craig PC. Memoirs of Sherlock Holmes through the eyes of an ophthalmologist. *Trans PA Acad Ophthalmol Otolaryngol* 1972;25(1):42-3.
5 Antiplatelet Trialists Collaboration. Collaborative overview of randomised trials of antiplatelet therapy. III: Reduction in venous thrombosis and pulmonary embolism by antiplatelet prophylaxis among surgical and medical patients. *BMJ* 1994;308:235-46.
6 Campling EA, Devlin HB, Hoile RW, Lunn JN. *Report of the national confidential enquiry into perioperative deaths 1992/1993*. London: NCEPOD, 1995.
7 Lowry JC. Thromboembolic disease and thromboprophylaxis in oral and maxillofacial surgery: experience and practice. *Br J Oral Maxillofac Surg* 1995;33:101-6.
8 Jones DR. Audit of attitudes to and use of postoperative thromboembolic prophylaxis in a regional health authority. *Ann R Coll Surg Engl* 1991;73:219-21.
9 Gent M, Roberts RS. A meta-analysis of the studies of dihydroergotamine plus heparin in the prophylaxis of deep vein thrombosis. *Chest* 1986;89(suppl 5):396-400S.
10 Collins R, Scrimgeour A, Yusuf S, Peto R. Reduction in fatal pulmonary embolism and venous thrombosis by perioperative administration of subcutaneous heparin: overview of results of randomized trials in general, orthopaedic and urologic surgery. *N Engl J Med* 1988;318:1162-73.
11 Daures JP, Schved JF, Momas I, Gril JC, Azoulay P, Gremy F. Meta-analysis on randomised trials comparing the results of low molecular weight heparins to those of fractioned heparins in the prevention of deep venous thrombosis [in French]. *Rev Epidemiol Sante Publique* 1989;37: 363-9.
12 Gallus AS, Cade JF, Mills KW, Murphy W. Apparent lack of synergism between heparin and dihydroergotamine in prevention of deep vein thrombosis after elective hip replacement: a randomised double-blind trial reported in conjunction with an overview of previous results. *Thromb Haemost* 1992;68:238-44.
13 Lassen MR, Borris LC, Christiansen HM, Schott P, Olsen AD, Sorensen JV, et al. Clinical trials with low molecular weight heparins in the prevention of postoperative thromboembolic complications: a meta-analysis. *Semin Thromb Hemost* 1991;17(suppl 3):284-90.
14 Greyer HG. Thromboprophylaxis with heparin and low molecular weight heparin in elective hip surgery: current status and perspectives. *Semin Thromb Hemost* 1991;17:336-42.
15 Nurmohamed MT, Rosendaal FR, Buller HR, Dekker E, Hommes DW, Vandenbroucke JP, et al. Low-molecular-weight-heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet* 1992;340:152-6.
16 Leizorovicz A, Haugh MC, Chapuis F-R, Samama MM, Boissel JP. Low molecular weight heparin in prevention of perioperative thrombosis. *BMJ* 1992;305:913-20.
17 Simonneau G, Leizorovicz A. Comparison of different methods of prevention of venous thrombosis in orthopaedics. *Ann Fr Anesth Reanim* 1992;11:354-62.
18 Anderson DR, O'Brien BJ, Levine MN, Roberts R, Wells PS, Hirsh J. Efficacy and cost of low-molecular-weight heparin compared with standard heparin for the prevention of deep vein thrombosis after total hip arthroplasty. *Ann Intern Med* 1993;119:1105-12.
19 Jorgensen LN, Wille-Jorgensen P, Hauch O. Prophylaxis of postoperative thromboembolism with low molecular weight heparins. *Br J Surg* 1993;80:689-704.
20 Mohr DN, Silverstein MD, Murtaugh PA, Harrison JM. Prophylactic agents for venous thrombosis in elective hip surgery: meta-analysis of studies using venographic assessment. *Arch Intern Med* 1993;153:2221-8.
21 Imperiale TF, Speroff T. A meta-analysis of methods to prevent venous thromboembolism following total hip replacement. *JAMA* 1994;271:1780-5.
22 Green D, Hirsh J, Heit J, Prins M, Davidson B, Lensing AW. Low molecular weight heparin: a critical analysis of clinical trials. *Pharmacol Rev* 1994;46:89-109.

23 Koch A, Bouges S, Ziegler S, Dinkel H, Daures JP, Victor N. Low molecular weight heparin and unfractionated heparin in thrombosis prophylaxis after major surgical intervention: update of previous meta-analysis. *Br J Surg* 1997;84:750-9.
24 Borris LC, Lassen MR, Jensen HP, Andersen BS, Poulsen KA. Perioperative thrombosis prophylaxis with low molecular weight heparins in elective hip surgery: clinical and economic considerations. *Int J Clin Pharmacol Ther* 1994;32:262-8.
25 O'Brien BJ, Anderson DR, Goeree R. Cost-effectiveness of enoxaparin versus warfarin prophylaxis against deep-vein thrombosis after total hip replacement. *Can Med Assoc J* 1994;150:1083-90.
26 Kearon C, Hirsch J. Starting prophylaxis for venous thromboembolism postoperatively. *Arch Intern Med* 1995;155:366-72.
27 Beyth RJ, Landefeld CS. Anticoagulants in older patients: a safety perspective. *Drugs Aging* 1995;6(1):45-54.
28 Borris LC, Lassen MR. A comparative review of the adverse effect profiles of heparins and heparinoids. *Drug Saf* 1995;12(1):26-31.
29 Murray DW, Britton AR, Bulstrode CJK. Thromboprophylaxis and death after total hip replacement. *J Bone Joint Surg* 1996;76B:863-70.
30 Andaz S, Shields A, Scurr JH, Coleridge-Smith PD. Role of low molecular weight heparins in the prevention and treatment of venous thromboembolism after surgery. *Phlebology* 1994;9:2-7.
31 Wells PS, Lensing AW, Hirsh J. Graduated compression stockings in the prevention of postoperative venous thromboembolism: a meta-analysis. *Arch Intern Med* 1994;154(1):67-72.
32 Clagett GP, Anderson FA Jr, Levine MN, Salzman EW, Wheeler HB. Prevention of venous thromboembolism. *Chest* 1992;102(suppl 4):391-407S.
33 Goldhaber SZ, Morpurgo M. Diagnosis, treatment, and prevention of pulmonary embolism. Report of the WHO/International Society and Federation of Cardiology Task Force. *JAMA* 1992;268:1727-33.
34 Thromboembolic Risk Factors (THRIFT) Consensus Group. Risk of and prophylaxis for venous thromboembolism in hospital patients. *BMJ* 1992;305:567-74.
35 Haas S. European consensus statement on the prevention of venous thromboembolism. *Blood Coag Fibrin* 1993;4(1):S5-8.
36 Skrabanek P. Nonsensus consensus. *Lancet* 1990;335:1446-7.
37 Conan Doyle A. Sherlock Holmes: the complete short stories. London: John Murray, 1928.
38 Egger M, Davey Smith G. Misleading meta-analysis. *BMJ* 1995;310:752-4.

(Accepted 26 August 1997)

Evaluating healthcare policies: the case of clinical audit

Joanne Lord, Peter Littlejohns

Health Care Evaluation Unit, Department of Public Health Sciences, St George's Hospital Medical School, London SW17 ORE

Joanne Lord, *health service research fellow*

Peter Littlejohns *director*

BMJ 1997;315:668–71

Clinicians are under increasing pressure to show that their services are effective and efficient. Some have recently suggested that policymakers should be subject to the same discipline.[1] Before the introduction of radical changes in the NHS in 1991 the government's decision not to pilot the proposals or evaluate them was widely criticised. One component of the reforms was a national programme to promote clinical audit by doctors, later extended to nurses and therapy professionals. This was generally welcomed, though dissenting voices questioned the underlying political motivations[2] and pointed to a lack of evidence on the value of clinical audit.[3-5]

Since then there have been repeated calls for the evaluation of audit,[6] and the public accounts committee has expressed concern at the failure to assess the overall cost effectiveness of the programme.[7] In fact, the NHS Executive has commissioned several evaluations of audit,[8 9] as well as monitoring progress through local and regional annual reports. However, it has not been possible to use scientifically rigorous methods to quantify the overall costs or benefits of national or local programmes of audit.[10] This paper describes the various approaches that have been tried (see box 1) and outlines the merits and disadvantages of each approach.

Summary points

Since the introduction of national programmes of clinical audit in Britain much effort has gone into evaluating them

Many observational studies, both quantitative and qualitative, have been conducted, but when these provide evidence of changes in clinical practice or outcomes it is not possible to attribute these to audit

No controlled trials of the introduction of whole programmes of audit into healthcare organisations have been conducted and it is too late to conduct one now

Several trials of selected audit interventions in Britain and elsewhere have been performed, but their results are not easily generalisable to mainstream audit activity

We still do not know, and will almost certainly never know, the scale of benefits or the true costs of the British national audit programmes

Evaluative research is worth while in indicating the types of audit activity and the types of audit organisation that are most likely to bring about change.

The classic model of evaluation

Evaluative research is essentially normative: it seeks to assess not just what is but also what ought to be. The classic rationalist model of evaluation consists of five basic steps: (*a*) identify the goals of the programme under evaluation; (*b*) translate the goals into measurable indicators; (*c*) measure indicators for the study group who have been exposed to the programme; (*d*) measure indicators for an equivalent group that has not been exposed to the programme; (*e*) compare the results for the experimental and control groups.[14]

Economic evaluation may be viewed as a special case of this means-ends analysis, where the prime goal is the maximisation of benefit from a given investment of resources. The method that most clearly exemplifies this model is the randomised controlled trial. The problems with trials are well known, but they remain the only way to guarantee freedom from bias and provide definitive answers to questions of effectiveness and cost effectiveness.

Randomised controlled trials of audit interventions

It is no longer possible to conduct a controlled trial to evaluate whole programmes of clinical audit in Britain:

Box 1
Methods used to evaluate clinical audit and examples

Experimental studies of specific audit projects
- Lomas et al's randomised controlled trial of "opinion leader education" and "audit with feedback" used to implement a caesarean section guideline[11]

Before-after studies of specific audit projects
- Lothian surgical audit: study of the impact of a surgical audit system on outcome indicators, clinical practice, and service organisation[12]

Quantitative observational studies of audit programmes
- Oxfordshire medical audit advisory group: annual review of primary care audit with appraisal of projects against criteria relating to progress around the audit cycle[13]

Qualitative studies of audit programmes
- CASPE review, national surveys of purchasers and providers followed by case studies at selected sites including interviews, meetings, and study of documents[8]

no provider organisations could serve as controls and any clinicians not participating in audit are highly selected. However, controlled trials of specific audit interventions have been conducted—for example, the Lomas study on the implementation of practice guidelines for caesarean section[11] and the north of England study of standard setting in general practice.[15]

The randomised controlled trial was designed to test individual clinical interventions and processes of care. Its extension to organisational technologies, such as clinical audit, is problematic[14] since it is difficult to define exactly what we mean by audit, and it is difficult to agree the goals of audit and to measure its impact.

Defining and controlling the audit intervention
As one might expect in an organisation as large and diverse as the NHS, approaches to audit, and quality in general, vary tremendously. There are many parallel initiatives, coming from a range of traditions, with varying philosophies and methods. In addition to this plurality of methods, audit is highly context dependent: it is contingent on personalities, relationships, professional and organisational structures, and processes.[16] Therefore trials of selected audit interventions in selected circumstances are not easily extrapolated to audit in general. For example, the "audit with feedback" intervention in the Lomas study is different from most clinical audit as routinely practised in Britain. Evaluation also requires some standardisation of the intervention under review, which is likely to change its nature. Even where investigators take care not to impose too tight a structure on the intervention, as in the north of England study, some artificiality is introduced.

Defining and measuring the impact of audit
Defining and operationalising the goals of clinical audit are surprisingly difficult. The ultimate aim should clearly be to improve the quality of patient care, but what do we mean by quality? Certainly good quality care must be clinically effective, but other factors, such as equity and respect for patients' autonomy, are also important. Furthermore, clinical audit seeks to improve patient care in various ways; as well as direct changes in clinical practice, there are potential indirect effects through professional education and team development. Turning abstract goals into measurable outcomes is also difficult. For individual projects it is often possible to define suitable outcome or process indicators—for example, the Lomas study used the proportion of women who had previously had a caesarean section who underwent a trial of labour. However, summarising such information for a range of projects is not straightforward and generic measures are insensitive.

Before-after studies of audit interventions
Several before-after studies of audit interventions have been conducted.[12] These may be less artificial than randomised controlled trials, but it is never really possible to isolate the effects due to the audit without a truly equivalent control group.

The methodological difficulties outlined above mean that classic methods of evaluation cannot tell us whether the requirement for routine clinical audit throughout the NHS or the creation of audit structures and processes has improved the quality of services overall, or whether similar or greater improvements could have been achieved by using resources in other ways. Systematic reviews of controlled trials and before-after studies of audit and other behaviourial change mechanisms show mixed results,[10 17 18] and results are not easily synthesised through meta-analyses because of heterogeneity of intervention and outcome measures. The Cochrane Collaboration on Effective Professional Practice is continuing work on this. However, tentative conclusions may be drawn about the types of intervention that are most effective—for example, "active" feedback, involving clinicians, appears to be more effective than "passive" feedback.

Alternative models of evaluation

The extension of classic models of decision making and evaluation into the realm of public policy has been criticised as both unrealistic and undesirable.[19] When high levels of complexity, uncertainty, and conflict exist it is often impossible to agree clear goals or to identify the best mechanisms for achieving these goals.

Several other approaches have been used to evaluate audit programmes, including various quantitative and qualitative observational studies. Qualitative approaches to evaluation differ fundamentally in the questions they seek to address—not just whether a programme is meeting its objectives but also why it is successful or unsuccessful. Qualitative, or "illuminative," approaches involve "intelligently using available situations, data, and methodology to produce best approximations to the otherwise unknowable relationships between cause and effect or between input and output."[14] They share four key characteristics (see box 2).

Quantitative observational studies of audit programmes
Given the difficulties in direct quantitative evaluation of clinical audit, a range of less direct methods have been used. Firstly, the level of audit activity has been assessed in terms of the numbers of clinicians participating,

Box 2
Four characteristics of qualitative evaluations

Subjective—measuring beliefs, attitudes, and perceptions about the subject of study and its value

Pluralist—searching for multiple perspectives from a range of interested parties

Eclectic—using a wide range of data sources, qualitative as well as quantitative, and using triangulation to test validity against more than one data source or perspective

Interpretive—investigating explanations for phenomena as well as, or instead of, hypothesis testing

time spent, or numbers of projects.[20] Secondly, the quality of audit projects has been appraised against process or quasi-outcome criteria, such as completion of the audit cycle.[13 21] Thirdly, attitudes towards audit and perceptions of the impact of audit have been assessed through surveys of provider staff.[22 23] These approaches are useful as they provide comparative information that can be used to improve audit. However, it is not necessarily the case that "more and better audits imply improved quality of care."[13] The philosophy of quality improvement rests on the assumption that staff participation and a sense of "ownership" are necessary to bring about change, which implies that clinicians must believe in clinical audit if it is to be effective, but the reverse is not necessarily true. In fact clinicians have been shown to have positive views about audit despite failing to complete the audit cycle.[21]

Qualitative evaluations of audit programmes

In addition to these pragmatic studies, many researchers have taken a qualitative approach.[8 9 16 24 25] For example, CASPE Research was commissioned by the Department of Health to evaluate the national programmes of audit.[8] They used a range of methods, including national surveys of purchasers and providers and detailed case studies. The CASPE evaluation team observed the functioning of audit at 29 selected providers, studying documents and conducting interviews. Progress was seen to vary considerably, with "very few" doing very well and a "substantial minority" very poorly. As Walshe points out, whether this observed range of performance represents success or failure is a matter of opinion.[8] From these observations CASPE proposed seven "critical success factors" for clinical audit programmes.

Just like classic quantitative methods of evaluation, qualitative approaches cannot answer our basic questions: Does audit work? Is it a good use of resources? They can, however, provide a rich picture of the perceived successes and failures of audit in practice and suggest some of the reasons for these successes and failures.

MARK HUDSON

Systems approaches

Finally, one approach to evaluation that has been little used in health services research is based on systems theory. The goal of systems engineering is to design systems to meet defined objectives while adapting to their environment.[26] Evaluation in this context entails comparing organisational structures and processes with some model of the "ideal adaptive organisation." As part of our study of clinical audit in South West Thames,[27] we used an approach based on one version of the systems approach called soft systems methodology.[19] Soft systems models may be both descriptive, a simplified representation of real world systems, and prescriptive, representing how systems should be. There are many possible versions of each type of model, so there are no right or wrong models.

We collected information through regional workshops, semistructured interviews, documents, and published reports. Alternative perspectives of how clinical audit does, could, or should work were articulated through conceptual models (see fig 1). Soft systems methods share the characteristics of other qualitative evaluations and have the same advantages and disadvantages.

Conclusions

"Research is a form of currency as varying interests negotiate a political resolution, but research is almost never definitive enough to resolve major issues on which strong political interests disagree."[28]

It is unlikely with the current techniques available that evaluative research will ever convince the sceptics of the value of clinical audit or disillusion the enthusiasts. A substantial body of research evaluating clinical audit now exists, but the debate continues.[29] Only the strongest level of evidence, the randomised controlled trial, could settle the controversy. But it is too late to conduct a trial of the introduction of local audit programmes, and trials of selected audit interventions cannot be extrapolated to the generality of routine audit. What evidence there is from experimental or quasi-experimental studies is equivocal.

Responding to questioning by the public accounts committee, the chief executive of the NHS in Scotland stated, "Since we're investing specific sums in a specific programme it does behove us to ask specific questions about what benefits are flowing and whether we can quantify these."[30] It also behoves us to be honest about what is achievable in terms of evaluating policy initiatives. We will never really know whether the national policy on clinical audit had a positive effect

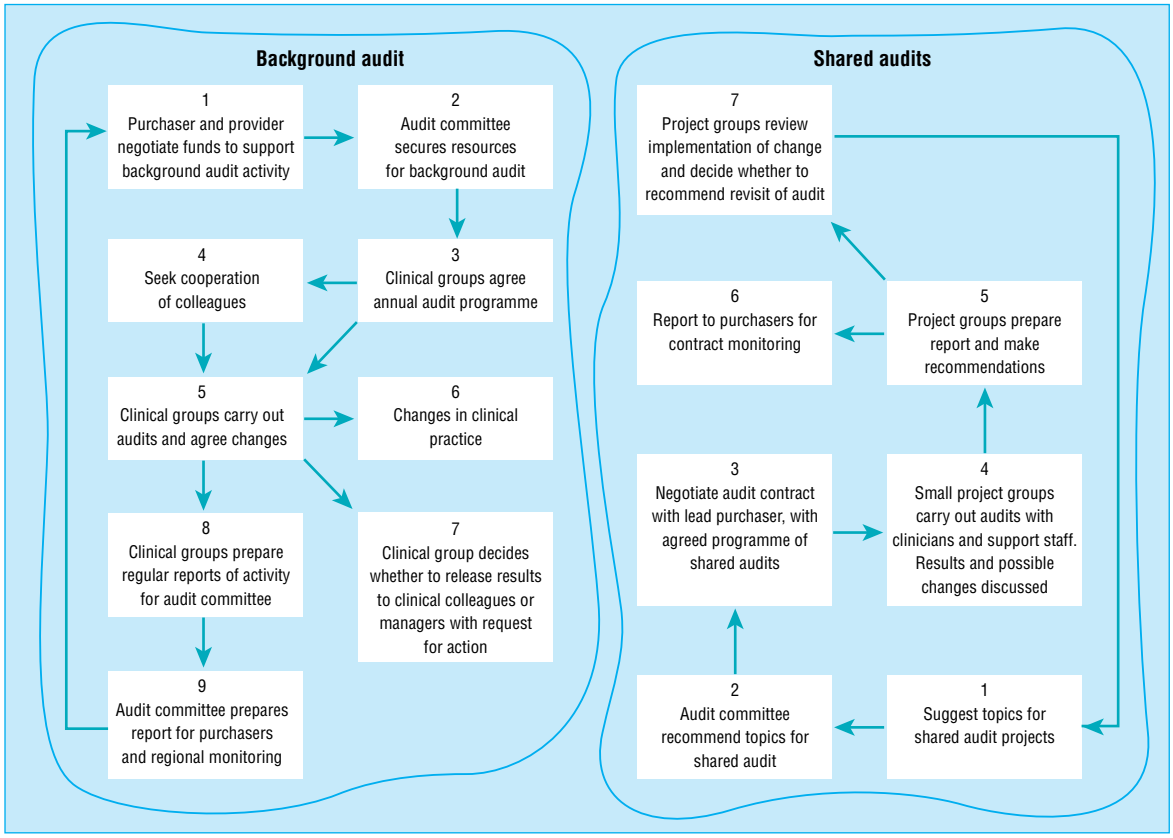

Fig 1 Example of a soft systems conceptual model. This split model illustrates one view of how a local programme of audit might be organised. It is designed to balance the conflicting needs for clinical ownership of audit and a supportive environment for professional development with the need to ensure that the concerns of others (including managers, purchasers, and patients) are addressed.

overall, or whether the money could have been better spent. Audit will always be an act of faith: a product of personal values, experience, professional loyalties, and anecdotal evidence.

This is not to say that evaluative research on organisational or policy issues is a waste of time; it is still useful to describe the impact of policy and explore the reasons for differing experiences. Both qualitative and quantitative approaches to evaluation can lead to practical prescriptions for improvement. There is already a wealth of information on clinical audit, and guidelines for effective audit have been developed.[18] These should reduce the number of unproductive, wasteful, and demoralising experiences of audit and increase the number of rewarding ones.

1 Ham C, Hunter DJ, Robinson R. Evidence based policymaking. *BMJ* 1995;310:71-2.
2 Sherwood T. Exitus auditus—no fun. *Lancet* 1992;340:37-8.
3 Maynard A. Case for auditing audit. *Health Services Journal* 1991;18 Jul:26.
4 Mooney G, Ryan M. Rethinking medical audit: the goal is efficiency. *J Epidemiol Community Health* 1992;46:180-3.
5 Buxton MJ. Achievements of audit in the NHS. *Quality in Health Care* 1994;3:S31-4.
6 Barton AG, Thomson RG, Bhopal RS. Clinical audit: more research is required. *J Epidemiol Community Health* 1995;49:445-7.
7 Committee of Public Accounts. *National health service executive clinical audit in England. 31st report.* London: HMSO,1996.
8 Walshe K, ed. *Evaluating clinical audit: past lessons, future directions.* London: Royal Society of Medicine Press, 1995.
9 Humphrey C, Berrow D. *Medical audit in primary care: a collation of evaluative projects 1991-93.* London: Department of Health, 1993.
10 Robinson MB. Evaluation of medical audit. *J Epidemiol Community Health* 1994;48:435-40.
11 Lomas J, Enkin M, Anderson GM, Hannah WJ, Vayda E, Singer J. Opinion leaders vs audit and feedback to implement practice guidelines: delivery after previous cesarean section. *JAMA* 1991;265:2202-7.
12 Gruer R, Gordon DS, Gunn AA, Ruckley CV. Audit of surgical audit. *Lancet* 1986;i:23-5.
13 Lawrence M, Griew K, Derry J, Anderson J, Humphreys J. Auditing audits: use and development of the Oxfordshire Medical Audit Advisory Group rating system. *BMJ* 1994;309:513-6.
14 Illsley R. *Professional or public health? Sociology in health and medicine. The Rock Carling Fellowship monograph.* London: Nuffield Provincial Hospitals Trust, 1980.
15 North of England Study of Standards and Performance in General Practice. Medical audit in general practice. II. Effect on health of patients with common childhood conditions. *BMJ* 1992;304:484-8.
16 Kerrison S, Packwood T, Buxton MJ. Monitoring medical audit. In: Robinson R, Le Grand J, eds. *Evaluating the NHS reforms.* London: King's Fund Institute, 1993;155-77.
17 Oxman AD, Thomson MA, Haynes RB, Davis DA. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *Can Med Assoc J* 1995;153:1423-31.
18 Dixon N. *Good practice in clinical audit—a summary of selected literature to support criteria for clinical audit.* London: National Centre for Clinical Audit, 1996.
19 Rosenhead J. *Rational analysis in a problematic world.* London: Wiley, 1989.
20 Webb SJ, Dowell AC, Heywood P. Survey of general practice audit in Leeds. *BMJ* 1991;302:390-2.
21 Eccles M, Deverill M, McColl E, Richardson H. A national survey of audit activity across the primary-secondary care interface. *Quality in Health Care* 1996;5:193-200.
22 Firth-Cozens J, Storer D. Registrars' and senior registrars' perceptions of their audit activities. *Quality in Health Care* 1992;1:161-4.
23 Lord J, Littlejohns P. Impact of hospital and community provider based clinical audit programmes: perceptions of doctors, nurses and other health professionals. *Int J Quality in Health Care* 1996;8:527-35.
24 Black N, Thompson E. Obstacles to medical audit: British doctors speak. *Soc Sci Med* 1993;36:849-56.
25 Thomson RG, Elcoat C, Pugh E. Clinical audit and the purchaser-provider interaction: different attitudes and expectations in the United Kingdom. *Quality in Health Care* 1996;5:97-103.
26 Gregory AJ, Jackson MC. Evaluation methodologies: a system for use. *J Operational Res Soc* 1992;43:19-28.
27 Lord J, Littlejohns P. Links between clinical audit and contracting systems. *Int J Health Care Quality Assurance* 1995;8:15-24.
28 Mechanic D. Social research in health and the American sociopolitical context: the changing fortunes of medical sociology. *Soc Sci Med* 1993;36:95-102.
29 Sellu D. Time to audit audit. *BMJ* 1996;312:128-9.
30 Committee of Public Accounts. *Auditing clinical care in Scotland.* London: HMSO, 1994.

(Accepted 2 February 1997)

How to read a paper

Papers that summarise other papers (systematic reviews and meta-analyses)

Trisha Greenhalgh

This is the ninth in a series of 10 articles introducing non-experts to finding medical articles and assessing their value

Unit for Evidence-Based Practice and Policy, Department of Primary Care and Population Sciences, University College London Medical School/ Royal Free Hospital School of Medicine, Whittington Hospital, London N19 5NF

Trisha Greenhalgh, *senior lecturer*

p.greenhalgh@ucl.ac.uk

BMJ 1997;315:672–5

Remember the essays you used to write as a student? You would browse through the indexes of books and journals until you came across a paragraph that looked relevant, and copied it out. If anything you found did not fit in with the theory you were proposing, you left it out. This, more or less, constitutes the methodology of the journalistic review—an overview of primary studies which have not been identified or analysed in a systematic (standardised and objective) way.

In contrast, a systematic review is an overview of primary studies which contains an explicit statement of objectives, materials, and methods and has been conducted according to explicit and reproducible methodology (fig 1).

Some advantages of the systematic review are given in box 1. When a systematic review is undertaken, not only must the search for relevant articles be thorough and objective, but the criteria used to reject articles as "flawed" must be explicit and independent of the results of those trials. The most enduring and useful systematic reviews, notably those undertaken by the Cochrane Collaboration, are regularly updated to incorporate new evidence.[2]

Many, if not most, medical review articles are still written in narrative or journalistic form. Professor Paul Knipschild has described how Nobel prize winning biochemist Linus Pauling used selective quotes from the medical literature to "prove" his theory that vitamin C helps you live longer and feel better.[3,4] When Knipschild and his colleagues searched the literature systematically for evidence for and against this hypothesis they found that, although one or two trials did strongly suggest that vitamin C could prevent the onset of the common cold, there were far more studies which did not show any beneficial effect.

Experts, who have been steeped in a subject for years and know what the answer "ought" to be, are less able to produce an objective review of the literature in their subject than non-experts.[5,6] This would be of little consequence if experts' opinions could be relied on to be congruent with the results of independent systematic reviews, but they cannot.[7]

Summary points

A systematic review is an overview of primary studies that used explicit and reproducible methods

A meta-analysis is a mathematical synthesis of the results of two or more primary studies that addressed the same hypothesis in the same way

Although meta-analysis can increase the precision of a result, it is important to ensure that the methods used for the review were valid and reliable

Evaluating systematic reviews

Question 1: Can you find an important clinical question which the review addressed?

The question addressed by a systematic review needs to be defined very precisely, since the reviewer must make a dichotomous (yes/no) decision as to whether each potentially relevant paper will be included or, alternatively, rejected as "irrelevant." Thus, for example, the clinical question "Do anticoagulants prevent strokes in patients with atrial fibrillation?" should be refined as an objective: "To assess the effectiveness and safety of warfarin-type anticoagulant therapy in secondary prevention (that is, following a previous stroke or transient ischaemic attack) in patients with non-rheumatic atrial fibrillation: comparison with placebo."[8]

Question 2: Was a thorough search done of the appropriate databases and were other potentially important sources explored?

Even the best Medline search will miss important papers, for which the reviewer must approach other sources.[9] Looking up references of references often yields useful articles not identified in the initial search,[10] and an exploration of "grey literature" (box 2) may be particularly important for subjects outside the medical

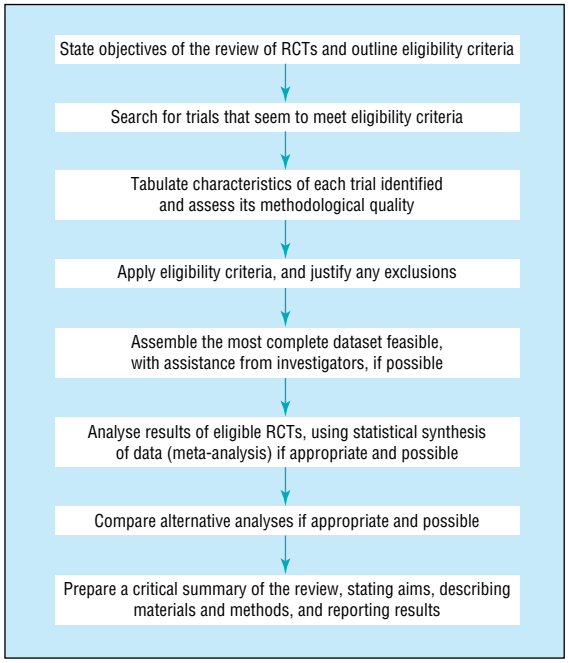

Fig 1 Methodology for a systematic review of randomised controlled trials[1]

Box 1
Advantages of systematic reviews[3]

- Explicit methods limit bias in identifying and rejecting studies
- Conclusions are more reliable and accurate because of methods used
- Large amounts of information can be assimilated quickly by healthcare providers, researchers, and policymakers
- Delay between research discoveries and implementation of effective diagnostic and therapeutic strategies may be reduced
- Results of different studies can be formally compared to establish generalisability of findings and consistency (lack of heterogeneity) of results
- Reasons for heterogeneity (inconsistency in results across studies) can be identified and new hypotheses generated about particular subgroups
- Quantitative systematic reviews (meta-analyses) increase the precision of the overall result

mainstream, such as physiotherapy or alternative medicine.[11] Finally, particularly where a statistical synthesis of results (meta-analysis) is contemplated, it may be necessary to write and ask the authors of the primary studies for raw data on individual patients which was never included in the published review.

Question 3: Was methodological quality assessed and the trials weighted accordingly?

One of the tasks of a systematic reviewer is to draw up a list of criteria, including both generic (common to all research studies) and particular (specific to the field) aspects of quality, against which to judge each trial (see box 3). However, care should be taken in developing such scores since there is no gold standard for the "true" methodological quality of a trial[12] and composite quality scores are often neither valid nor reliable in practice.[13 14] The various Cochrane collaborative review groups are developing topic-specific methodology for assigning quality scores to research studies.[15]

Question 4: How sensitive are the results to the way the review has been done?

Carl Counsell and colleagues "proved" (in the Christmas 1994 issue of the *BMJ*) an entirely spurious relationship between the result of shaking a dice and the outcome of an acute stroke.[16] They reported a series of artificial dice rolling experiments in which red, white, and green dice represented different therapies for acute stroke. Overall, the "trials" showed no significant benefit from the three therapies. However, the simulation of a number of perfectly plausible events in the process of meta-analysis—such as the exclusion of several of the "negative" trials through publication bias, a subgroup analysis which excluded data on red dice therapy (since, on looking back at the results, red dice appeared to be harmful), and other, essentially arbitrary, exclusions on the grounds of "methodological quality"—led to an apparently highly significant benefit of "dice therapy" in acute stroke.

If these simulated results pertained to a genuine medical controversy, how would you spot these subtle biases? You need to work through the "what ifs". What if the authors of the systematic review had changed the inclusion criteria? What if they had excluded unpublished studies? What if their "quality weightings" had been assigned differently? What if trials of lower methodological quality had been included (or excluded)? What if all the patients unaccounted for in a trial were assumed to have died (or been cured)?

An exploration of what ifs is known as a sensitivity analysis. If you find that fiddling with the data in various ways makes little or no difference to the review's overall results, you can assume that the review's conclusions are relatively robust. If, however, the key findings disappear when any of the what ifs changes,

Box 2
Checklist of data sources for a systematic review

- Medline database
- Cochrane controlled clinical trials register
- Other medical and paramedical databases
- Foreign language literature
- "Grey literature" (theses, internal reports, non-peer reviewed journals, pharmaceutical industry files)
- References (and references of references, etc) listed in primary sources
- Other unpublished sources known to experts in the field (seek by personal communication)
- Raw data from published trials (seek by personal communication)

Box 3
Assigning weight to trials in a systematic review

Each trial should be evaluated in terms of its:
- Methodological quality—the extent to which the design and conduct are likely to have prevented systematic errors (bias)
- Precision—a measure of the likelihood of random errors (usually depicted as the width of the confidence interval around the result)
- External validity—the extent to which the results are generalisable or applicable to a particular target population

PETER BROWN

the conclusions should be expressed far more cautiously and you should hesitate before changing your practice in the light of them.

Question 5: Have the numerical results been interpreted with common sense and due regard to the broader aspects of the problem?

Any numerical result, however precise, accurate, "significant," or otherwise incontrovertible, must be placed in the context of the painfully simple and often frustratingly general question which the review addressed. The clinician must decide how (if at all) this numerical result, whether significant or not, should influence the care of an individual patient. A particularly important feature to consider when undertaking or appraising a systematic review is the external validity or relevance of the trials that are included.

Meta-analysis for the non-statistician

A good meta-analysis is often easier for the non-statistician to understand than the stack of primary research papers from which it was derived. In addition to synthesising the numerical data, part of the meta-analyst's job is to tabulate relevant information on the inclusion criteria, sample size, baseline patient characteristics, withdrawal rate, and results of primary and secondary end points of all the studies included. Although such tables are often visually daunting, they save you having to plough through the methods sections of each paper and compare one author's tabulated results with another author's pie chart or histogram.

These days, the results of meta-analyses tend to be presented in a fairly standard form, such as is produced by the computer software MetaView. Figure 2 is a pictorial representation (colloquially known as a "forest plot") of the pooled odds ratios of eight randomised controlled trials which each compared coronary artery bypass grafting with percutaneous coronary angioplasty in the treatment of severe angina.[17] The primary (main) outcome in this meta-analysis was death or heart attack within one year.

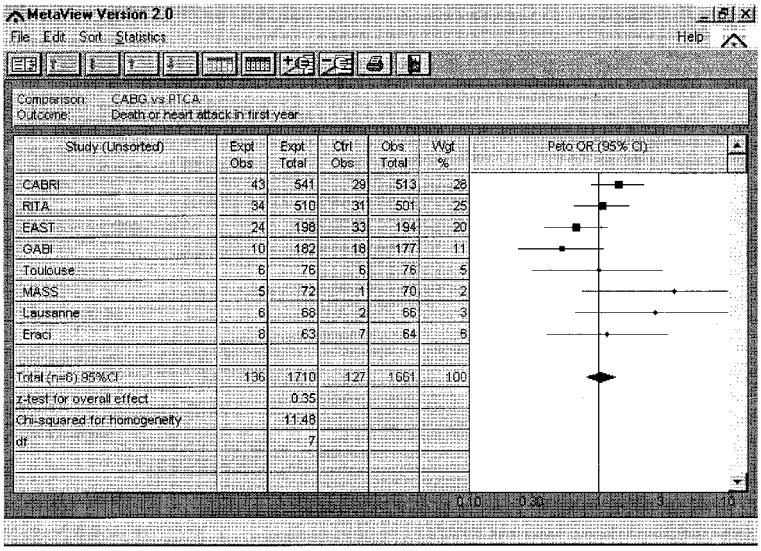

Fig 2 Pooled odds ratios of eight randomised controlled trials of coronary artery bypass grafting against percutaneous coronary angioplasty, shown in MetaView format. Reproduced with authors' permission[17]

The horizontal line corresponding to each of the eight trials shows the relative risk of death or heart attack at one year in patients randomised to coronary angioplasty compared to patients randomised to bypass surgery. The "blob" in the middle of each line is the point estimate of the difference between the groups (the best single estimate of the benefit in lives saved by offering bypass surgery rather than coronary angioplasty), and the width of the line represents the 95% confidence interval of this estimate. The black line down the middle of the picture is known as the "line of no effect," and in this case is associated with a relative risk of 1.0.

If the confidence interval of the result (the horizontal line) crosses the line of no effect (the vertical line), that can mean either that there is no significant difference between the treatments or that the sample size was too small to allow us to be confident where the true result lies. The various individual studies give point estimates of the relative risk of coronary angioplasty compared with bypass surgery of between about 0.5 and 5.0, and the confidence intervals of some studies are so wide that they do not even fit on the graph. Now look at the tiny diamond below all the horizontal lines. This represents the pooled data from all eight trials (overall relative risk of coronary angioplasty compared with bypass surgery = 1.08), with a new, much narrower, confidence interval of this relative risk (0.79 to 1.50). Since the diamond firmly overlaps the line of no effect, we can say that there is probably little to choose between the two treatments in terms of the primary end point (death or heart attack in the first year). Now, in this example, every one of the eight trials also suggested a non-significant effect, but in none of them was the sample size large enough for us to be confident in that negative result.

Note, however, that this neat little diamond does not mean that you might as well offer coronary angioplasty rather than bypass surgery to every patient with angina. It has a much more limited meaning—that the average patient in the trials presented in this meta-analysis is equally likely to have met the primary outcome (death or myocardial infarction within a year), whichever of these two treatments they were randomised to receive. If you read the paper by Pocock and colleagues[17] you would find important differences in the groups in terms of prevalence of angina and requirement for further operative intervention after the initial procedure.

Explaining heterogeneity

In the language of meta-analysis, homogeneity means that the results of each individual trial are mathematically compatible with the results of any of the others. Homogeneity can be estimated at a glance once the trial results have been presented in the format illustrated in figures 2 and 3. In figure 2 the lower confidence limit of every trial is below the upper confidence limit of all the others (that is, the horizontal lines all overlap to some extent). Statistically speaking, the trials are homogeneous. Conversely, in figure 3 some lines do not overlap at all. These trials may be said to be heterogeneous.

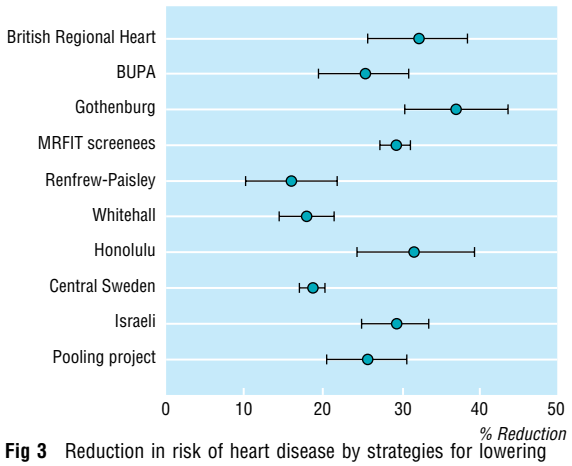

Fig 3 Reduction in risk of heart disease by strategies for lowering cholesterol. Reproduced with permission from Chalmers and Altman[18]

The definitive test for heterogeneity involves a slightly more sophisticated statistical manoeuvre than holding a ruler up against the forest plot. The one most commonly used is a variant of the χ^2 (chi square) test, since the question addressed is whether there is greater variation between the results of the trials than is compatible with the play of chance. Thompson[18] offers the following rule of thumb: a χ^2 statistic has, on average, a value equal to its degrees of freedom (in this case, the number of trials in the meta-analysis minus one), so a χ^2 of 7.0 for a set of eight trials would provide no evidence of statistical heterogeneity. Note that showing statistical heterogeneity is a mathematical exercise and is the job of the statistician, but explaining this heterogeneity (looking for, and accounting for, clinical heterogeneity) is an interpretive exercise and requires imagination, common sense, and hands-on clinical or research experience.

Figure 3 shows the results of ten trials of cholesterol lowering strategies. The results are expressed as the percentage reduction in risk of heart disease associated with each reduction of 0.6 mmol/l in serum cholesterol concentration. From the horizontal lines which represent the 95% confidence intervals of each result it is clear, even without knowing the χ^2 statistic of 127, that the trials are highly heterogeneous. Correcting the data for the age of the trial subjects reduced this value to 45. In other words, much of the "incompatibility" in the results of these trials can be explained by the fact that embarking on a strategy which successfully reduces your cholesterol level will be substantially more likely to prevent a heart attack if you are 45 than if you are 85.

Clinical heterogeneity, essentially, is the grievance of Professor Hans Eysenck, who has constructed a vigorous and entertaining critique of the science of meta-analysis.[19] In a world of lumpers and splitters, Eysenck is a splitter, and it offends his sense of the qualitative and the particular to combine the results of studies which were done on different populations in different places at different times and for different reasons.

Eysenck's reservations about meta-analysis are borne out in the infamously discredited meta-analysis which showed (wrongly) that giving intravenous magnesium to people who had had heart attacks was beneficial. A subsequent megatrial involving 58 000 patients (ISIS-4) failed to find any benefit, and the meta-analysts' misleading conclusions were subsequently explained in terms of publication bias, methodological weaknesses in the smaller trials, and clinical heterogeneity.[20 21]

Thanks to Professor Iain Chalmers for advice on this chapter.

The articles in this series are excerpts from *How to read a paper: the basics of evidence based medicine*. The book includes chapters on searching the literature and implementing evidence based findings. It can be ordered from the BMJ Publishing Group: tel 0171 383 6185/6245; fax 0171 383 6662. Price £13.95 UK members, £14.95 non-members.

1 *The Cochrane Centre*. Cochrane Collaboration Handbook [updated 9 December 1996]. The Cochrane Collaboration; issue 1. Oxford: Update Software, 1997.
2 Bero L, Rennie D. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995:274:1935-8.
3 Chalmers I, Altman DG, eds. *Systematic reviews*. London: BMJ Publishing Group, 1995.
4 Pauling L. *How to live longer and feel better*. New York: Freeman, 1986.
5 Oxman AD, Guyatt GH. The science of reviewing research. *Ann NY Acad Sci* 1993; 703: 125-31.
6 Mulrow C. The medical review article: state of the science. *Ann Intern Med* 1987;106: 485-8.
7 Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomised controlled trials and recommendations of clinical experts. *JAMA* 1992;268:240-8.
8 Koudstaal P. Secondary prevention following stroke or TIA in patients with non-rheumatic atrial fibrillation: anticoagulant therapy versus control. *Cochrane Database of Systematic Reviews*. Oxford: Cochrane Collaboration, 1995. (Updated 14 February 1995.)
9 Greenhalgh T. Searching the literature. In: *How to read a paper*. London: BMJ Publishing Group, 1997:13-33.
10 Knipschild P. Some examples of systematic reviews. In: Chalmers I, Altman DG. *Systematic reviews*. London: BMJ Publishing Group, 1995:9-16.
11 Knipschild P. Searching for alternatives: loser pays. Lancet 1993; 341: 1135-6.
12 Oxman A, ed. Preparing and maintaining systematic reviews. In: *Cochrane Collaboration handbook, section VI*. Oxford: Cochrane Collaboration, 1995. (Updated 14 July 1995.)
13 Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Controlled Clin Trials* 1990;11:339-52.
14 Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. *Int J Health Technol Assess* 1996;12:195-208.
15 Garner P, Hetherington J. Establishing and supporting collaborative review groups. In: *Cochrane Collaboration handbook, section II*. Oxford: Cochrane Collaboration, 1995 (Updated 14 July 1995.)
16 Counsell CE, Clarke MJ, Slattery J, Sandercock PAG. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? *BMJ* 1994;309:1677-81.
17 Pocock SJ, Henderson RA, Rickards AF, Hampton JR, Sing SB III, Hamm CW, et al. Meta-analysis of randomised trials comparing coronary angioplasty with bypass surgery. *Lancet* 1995;346:1184-9.
18 Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. In: Chalmers I, Altman DG. *Systematic reviews*. London, BMJ Publishing Group, 1995:48-63.
19 Eysenck HJ. Problems with meta-analysis. In: Chalmers I, Altman DG. *Systematic reviews*. London: BMJ Publishing Group, 1995:64-74.
20 Magnesium, myocardial infarction, meta-analysis and mega-trials. *Drug Ther Bull* 1995;33:25-7.
21 Egger M, Davey Smith G. Misleading meta-analysis: lessons from "an effective, safe, simple" intervention that wasn't. *BMJ* 1995;310:752-4.

Correction

Statistics for the non-statistician. I: Different types of data need different tests

An author's error appeared in this article by Trisha Greenhalgh (9 August, pp 364-6). In table 1, the χ^2 test is listed as a parametric test. In fact, both the χ^2 test and Fisher's exact test are non-parametric.