Editorial

Assessment of evidence versus consensus or prejudice

Assessing evidence with the minimum of observer bias is central to scientific enquiry and hypothesis testing—so central it seems quite unnecessary to remind ourselves of this. Yet many doctors are principally scientists of the “classification” school, perhaps partly as a consequence of the extent of medical knowledge, the length of medical training, and the amount of information to be learnt by rote; and many researching scientists operate more as “technicians”, perhaps partly as a consequence of academic competition, pressure to publish, and pressure to maintain grant income. With an ever increasing medical and scientific literature and an ever increasing complexity of medical research, it is important that both medical research scientists and leading clinicians are able to assess the evidence and appraise the literature critically and to minimise personal bias while doing so.

Medical schools are becoming increasingly conscious of this need and courses for undergraduate medical students and short courses for registrars, seeking guidance in research training, are designed specifically to address the issue of “critical appraisal of published medical/scientific evidence”.^{1–4} These developments and courses are generally based in epidemiology (or social medicine, community medicine, or public health) departments. Techniques of “critical appraisal” are useful in their own right but what is possibly being missed in these developments is discussion of the fundamental principles underlying the need for these techniques; that is, to minimise as far as possible personal biases of the hypothesis generators and testers, observers, experimenters, and readers.

Historians and philosophers of science have taught how scientific knowledge advances often not as a smooth logical progression but in quantum leaps from one established paradigm to another and furthermore that it often takes many attempts to leave the security of one consensus theory, truth, or fact before it is suddenly appreciated that the new theory is truer and fits the observations better than the old.^{5–7} With the wisdom of hindsight many outspoken advocates of the “flat earth society” became equally vociferous supporters of the planetary theory of the solar system. The challenge for the practising scientist is to identify the elements of data that do not fit the flat earth theory, when that is the predominant and socially accepted reigning theory, and to proffer alternative explanations, hypotheses, or theories.

With this background it is salutary to consider a recent “case study”, which is in itself an important case since it relates to control of one of the primary risk factors of the leading cause of death in the western world, and correct interpretation of the literature is important for public health as presently practised. The association between serum cholesterol and mortality from heart disease observed in cross sectional, case-control, and cohort studies led to major population wide experimental evaluation of cholesterol lowering regimens. A number (more than 20) of important primary prevention trials have been undertaken, utilising diet modification or cholesterol lowering medication. The accumulating evidence has been and no doubt will continue to be reviewed.^{8–10}

Reviewing used to be (mostly) an “art”, practised by elder statesmen of the subject, drawing on their accumulated experience and wisdom, sometimes possibly with statistical and other specialist advice in the background. In recent years this “art” had been given a more scientific basis by the introduction of the statistical overview or meta-analysis.^{11 12} For those unfamiliar with the statistical overview, the technique may be described simply as selecting from published reports trials testing a common hypothesis and satisfying certain criteria of patient selection (or exclusion), intervention (drug or operation), completeness of recording and follow up, and measurement of outcome, and then calculating the average effect of the selected trials.

A recent review of cholesterol lowering and mortality was such a statistical overview (or meta-analysis) of six selected primary trials. It appeared under the familiar format of a scientific paper, with introduction, methods, results, and discussion.¹³ This in itself could be misleading for cursory or skimming readers, unfamiliar with the recent developments in statistical overviews, since the article, despite its title, could be mistaken for a new and very large primary trial. The average finding of the trials included in that statistical overview was that there was no benefit of lowering cholesterol with respect to total mortality, although there may have been a transfer of cause of death from heart disease to other causes: the review highlighted suicide and violence. The review led, not perhaps surprisingly, to a somewhat excited correspondence (*BMJ* 15 September and 6 October, 1990). The range of reactions was wide, as the authors observed in their response: “We are said to have both understated and overstated the adverse effects”.

While discussion is essential for the advancement of science and good debate is an important part of that essential discussion, some of the arguments used in this correspondence are disquieting for medical science and perhaps particularly so for epidemiology as a discipline, if epidemiologists wish to be the “guardians of rational appraisal of medical/scientific evidence”.^{1–4} It is not the intention here to cap the extensive reviewing of the cholesterol story, but to draw attention to some illogicalities and irrationalities in arguments presented and the powerful influence of the reigning consensus expert view. It was after all a “case history” and not in any way unique, since debate or argument surrounds many reports and publications. We are concerned here about general issues of the rationality of scientific debate, the assessment of scientific evidence, and the development of scientific knowledge.

Medical scientists and epidemiologists in particular teach the next generation of medical scientists and doctors that the randomised controlled trial sits at the head of a “hierarchy of evidence”³ and is the best demonstration of cause and effect, that bigger trials have more power than smaller trials, and that many trials are more representative than single trials. Yet as this “case study” shows, medical scientists, including epidemiologists, when faced by the pooled results of many large randomised trials that appear to contradict their “prior belief model” or the “consensus view” may abandon their own

teaching and recommend instead the weaker evidence of observational studies and even weaker and more subjective observations of clinical experience. Such is the power of "scientific consensus", conformity within the scientific community, or "personal prejudice". There are acknowledged problems with the statistical overview, which have been discussed more fully elsewhere.^{14 15} The principal problem arises through the selection of evidence post hoc, which clearly contradicts a fundamental scientific principle of postulating hypotheses and then designing experiments to test the hypotheses. It is also true that this particular statistical overview of cholesterol lowering trials was highly selective (six of more than 20 trials). Nevertheless it is apparent that emotion contributed to the correspondence that emanated from this case.

Research scientists and epidemiologists working in the abstract and with the population approach, are very familiar with the concept of bias and the possible effects of bias on scientific observation. We may not, however, fully appreciate the "inevitability of prejudice"¹⁶ and that we may be as humanly subject to prejudice as the non-scientifically trained. Scientists should be sufficiently self critical to be aware that they may enter a debate with an established prior belief, bias, or prejudice, cling to those beliefs for the security and respectability which accompany conformity, and seek to maintain established medical or scientific paradigms by selecting evidence that supports those beliefs or paradigms. It is widely recognised that it takes a lot to dislodge established consensus views or explanatory models and to replace the old with new explanatory models^{5 6} although scientific philosophers have suggested that correction is as important as accumulation for the advancement of scientific knowledge.^{7 17} The history of science is replete with examples.

To return briefly to the cholesterol story before citing other epidemiological examples, it is perhaps time for revision of the current consensus paradigm (paraphrased very simply as) "cholesterol causes heart disease and heart disease leads to death, therefore cholesterol leads to death and consequently lowering cholesterol averts death", in the light of the pooled evidence of large randomised controlled trials. Another example in cardiovascular epidemiology is the debate surrounding exercise based rehabilitation following myocardial infarction. Briefly, the WHO European rehabilitation trials group reported no net life saving benefit (one centre better, one worse, others non-significant).¹⁸ However, those trials were evaluating multifactorial programmes and more recently statistical overviews of selected trials (from the WHO group and from elsewhere) in which exercise was believed (by the post hoc overviews) to be the active ingredient have suggested a 20% net reduction in mortality.¹⁹ The clinical consensus remains somewhat sceptical: editorials suggests that rehabilitation programmes are unnecessary,²⁰ practitioners wonder whether exercise may not be the "active ingredient", and researchers look to evaluate other modes, particularly those using the psychological approach.²¹ There are similarities with the cholesterol primary prevention trials: the evidence from the statistical overview of 22 randomised controlled trials is being refuted because the "mechanism is not fully understood" or because of prior prejudice?

Epidemiologists will recognise that there are examples of controversy in many areas of medicine and public health and that several have attempted to address the issues of unbiasing the experimenter and the appraiser of evidence.^{22 23} The statistical overview (or meta-analysis) is a useful tool and can provide a larger less biased pool of evidence, but as illustrated by the above examples it should be appreciated that it is no more than a tool and does not overcome the problem of publication bias. Advocates for statistical overview recommend including results from unpublished studies but those are often very difficult to identify and many "negative" studies are never even completed. Publication bias could be

reduced if papers were refereed on the basis of hypothesis, methods, and sample size without results.^{14 22} A further extension of less biased information could be obtained if studies were "registered" at outset (when funded) and researchers were obliged to provide some results (or report) to a pooling project, even if the study turned out very negatively and was aborted. Much has been achieved in these directions in the area of perinatal trials by the Oxford database.²³ There are many other fields of application of the clinical trial where bias reduction could greatly benefit the accumulation of scientific knowledge and perhaps the revision of scientific consensus. There is also the even wider field in which to attempt bias reduction in knowledge arising from observational (cohort and case-control) studies.

In conclusion, researchers, particularly epidemiologists, might reflect more deeply on their own teaching with respect to the "hierarchy of evidence",¹⁻⁴ the effects of observer bias, publication bias and reader bias, and their own possible prejudices with respect to established beliefs,^{5 7 16-18} when contributing to scientific argument. If we are agreed that population based studies are preferable to clinical experiences for associational inference, that experimental trials are better at explaining cause than observational studies, and that many trials are more convincing than single trials, in general we should accept the same hierarchy when reviewing evidence apparently contradicting the established consensus (or our own beliefs) as when reviewing evidence supporting them. We could usefully remember the empirical discovery of both the aetiology of and the treatment for scurvy more than a century before vitamins were identified and mechanisms explained.²⁴ McIntyre and Popper, in calling for a critical attitude in medical practice and supporting the case for audit, recommended both self criticism and critical appraisal of the established best tested theories, as essential to advancement of scientific knowledge.¹⁷

ROBERT WEST

- Gehlbach SH. *Interpreting the medical literature*. Lexington: Collamore Press, 1982.
- Sackett DL, Haynes RB, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. Boston: Little Brown, 1985.
- Elwood JM. *Causal relationships in medicine*. London: Oxford Medical, 1988.
- Campbell MJ, Machin D. *Medical statistics: a common sense approach*. London: John Wiley, 1990.
- Popper K. *The logic of scientific discovery*. London: Hutchinson, 1959.
- Kuhn TS. *Structure of scientific revolutions*. Chicago: University of Chicago Press, 1962.
- Popper K. *Objective knowledge: an evolutionary approach*. London: Oxford University Press, 1972.
- Mitchell JRA. What constitutes evidence on the dietary prevention of coronary heart disease? Cosy beliefs or hard facts. *Int J Cardiol* 1984; 5: 287-98.
- Consensus conference. Lowering blood cholesterol to prevent heart disease. *JAMA* 1985; 253: 2080-6.
- British Cardiac Society. Working group on coronary prevention: conclusions and recommendations. *Br Heart J* 1987; 57: 188-9.
- Peto R. Clinical trial methodology. *Biomedicine* (special issue) 1978; 28: 24-36.
- Pocock SJ. *Clinical trials, a practical approach*. Chichester: John Wiley, 1983.
- Muldoon MF, Manuch SB, Matthews KA. Lowering cholesterol concentrations and mortality, a quantitative review of primary prevention trials. *BMJ* 1990; 301: 309-14.
- Begg CB, Berlin JA. Publication bias, a problem in interpreting medical data. *J R Stat Soc* 1988; 151: 419-63.
- Spector TD, Thompson SG. Potential and limitations of meta analysis. *J Epidemiol Community Health* 1991; 45: 89-92.
- West R. The inevitability of prejudice. *Ethics* 1952; 62: 205-9.
- McIntyre N, Popper K. Critical attitudes in medicine: need for a new ethic. *BMJ* 1983; 287: 1919-23.
- Kallio V, Cay EL, eds. *Rehabilitation after myocardial infarction*. Public Health in Europe. 24. Copenhagen: World Health Organization, 1985.
- O'Connor GT, Buring JE, Yusuf S, et al. Overview of randomised trials of rehabilitation after myocardial infarction. *Circulation* 1989; 80: 234-44.
- Lipkin D. Is cardiac rehabilitation necessary? *Br Heart J* 1991; 65: 237-8.
- Jones DA, West RR. Multicentre randomised controlled trial of rehabilitation following myocardial infarction. *Br Heart Foundation annual report*, 1991.
- Walster GW, Cleary TA. Proposal for new editorial policy in social sciences. *Am Stat* 1970; 24: 16-19.
- Chalmers I, Enkin M, Keirse MJ, eds. *Effective care in pregnancy and childbirth*. Oxford: Oxford University Press, 1989.
- Lind J. *Treatise on the scurvy*. Edinburgh: Sands, Murray and Cochrane, 1753.