findings from a literature review with an analysis of routine observational data generated mainly from databases held by Medicare, insurance companies, and hospitals. Though many of the reviews and guidelines produced by the Agency for Health Care Policy and Research are excellent, the central element of the outcomes research teams' strategies is fundamentally flawed,[3] and this book reveals the weak scientific foundations of the programme.

One of the most important methods for assessing whether treatments really do more good than harm is the randomised controlled trial, made famous by Bradford Hill half a century ago. As long as they are sufficiently large, such trials are valid methods of evaluating interventions because if patients are randomised to alternative treatment groups differences in outcome between the groups can be more confidently attributed to the difference in treatments received.[4]

In non-randomised observational studies, however, patients receiving different treatments may differ systematically with respect to any number of known and unknown factors that affect prognosis. These include the severity of the main and accompanying disease, clinical setting, and clinician. Although statistical adjustments may be made in an attempt to exclude the effects of these confounders (and thus isolate any differences due solely to the treatment), this assumes both a complete knowledge of the confounding variables and their comprehensive and accurate measurement. Neither is likely to be possible, and at least a moderate bias will remain.[5] As most common treatments that interest us will probably have only moderately sized effects (though with a large absolute benefit in large populations) the ability to exclude even moderate effects of confounding is vital.

Despite its considerable cost the American programme does not seem to have made a substantial contribution to our knowledge of effectiveness in any field through the analysis of observational data. This is in stark contrast to the contribution made by large scale simple randomised controlled trials and properly conducted overviews of such trials. For example, in this book Peto and colleagues describe four major examples—the survival gains incurred with thrombolysis for acute myocardial infarction, use of aspirin for people at high risk of thrombotic events, adjuvant treatment for early breast cancer, and the evidence for a lack of benefit with infusions of magnesium in suspected myocardial infarction. Some of the overviews in the Cochrane Pregnancy and Childbirth database have contributed to the quality of care for women using maternity services.[6]

In Britain the lack of routinely collected data on health care process and outcome, though a national disgrace, has largely protected us from the mirage of quick and easy answers from analyses of databases. The government's misuse of the limited data that it collects, such as school examination results and hospital waiting times, to produce crude and meaningless league tables for comparing institutions should sound a cautionary note.[7]

No short cuts exist for obtaining reliable information on effectiveness. Large multicentre simple randomised controlled trials and meta-analyses of trails can answer reliably a wide range of questions about the effectiveness of treatments. The challenges now are to design trials that provide answers to more clinically relevant questions, such as which patients stand to benefit most.[8] We need to ensure that meta-analysis is used appropriately and incorporates analysis of adequate sensitivity[9]; we need to include the outcome measures that matter to patients; and we need to work harder to get the results of research into practice.

Last year the patient outcomes research teams seemed secure. In the face of the sort of criticism summarised in this book, however, a major shift was recently announced for the second phase, emphasising other methods such as clinical trials. This is a welcome change: diverting money from relatively cost effective trials to uninformative analyses of databases may do more harm than good.

TREVOR A SHELDON
Director

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

1 Warren KS, Mosteller F, eds. Doing more good than harm: the evaluation of health care interventions. *Ann NY Acad Sci* 1993;**703**:1-341.
2 Anderson C. Measuring what works in health care. *Science* 1994;**263**:1080-1.
3 Frater C, Sheldon TA. The outcomes movement in the USA and the UK. In: Drummond MF, Maynard A, eds. *Purchasing and providing cost-effective health care.* Edinburgh: Churchill Livingstone, 1993:49-65.
4 Hill AB. Observation and experiment. *N Engl J Med* 1953;**248**:995-1001.
5 Jencks SF. Issues in the use of large databases for effectiveness research. In: Heithoff KA, Lohr KN, eds. *Effectiveness and outcomes in health care.* Washington: National Academy Press, 1990;94-104.
6 Lilford R. The Cochrane Pregnancy and Childbirth database (book review). *BMJ* 1994;**308**:1448.
7 Dillner L. League tables criticised as misleading. *BMJ* 1994;**309**:9-10.
8 Davey Smith G, Egger M. Who benefits from medical interventions? *BMJ* 1994;**308**:72-4.
9 Egger M, Davey Smith G. Magnetism and myocardial infarction. *Lancet* 1994;**343**:1285

Promoting research into peer review

An invitation to join in

One of the jobs of the editor of the *BMJ* is to sit on the editorial boards of the 18 special journals owned by the BMJ Publishing Group. At the meetings of these boards many questions are asked about the peer review process. What can be done to raise the standard of reviewing? How many reviewers should be used? How should they be selected? Should reviewers be blinded to the names of the authors? Or, in contrast, should the traditional anonymity of reviewers be abolished? Why are reviewers anonymous anyway? Should authors be encouraged to suggest reviewers for their own papers? Will this corrupt the peer review process? Should statistical reviewers be used? Should they be used for all papers, and should they be used before or after clinical reviewers? The same questions arise at most meetings, and no doubt they arise at the editorial boards of all the many thousands of biomedical journals published around the world.

Many of the people on the boards have a long experience of reviewing papers, and many of them have strong opinions on the questions that arise. Many, indeed, are confident that they know the answers. Yet the members of the boards—most of whom are doctors and clinical researchers—usually make their statements without producing any systematic evidence whatsoever. People who increasingly in their clinical lives make decisions on the treatment of patients on the basis of scientifically

sound, peer reviewed, published evidence seem content to make decisions on peer review processes without any such evidence. The contrast is striking. The problem arises partly because there is little high quality research on peer review and partly because members of editorial boards are largely unaware of the evidence that does exist.

Yet peer review of papers and grant applications is a hugely expensive process that consumes large amounts of academic time and has great consequences for the development of medical science and for the careers of authors and researchers.[1-3] Furthermore, much of what is published in peer reviewed journals is of very low quality.[4,5] We must do better with developing our scientific understanding of peer review, and the publication this week of the second theme issue of *JAMA* devoted to the subject is an important step forward.[6]

Although many members of editorial boards and many editors may be unaware of the results of research into peer review, the call to begin such research was made more than a decade ago by pioneers like Stephen Lock,[1] my predecessor as editor of the *BMJ*; Drummond Rennie, the deputy editor of *JAMA*[7]; and John Bailar, the statistical adviser to the *New England Journal of Medicine*.[8] The theme issue of *JAMA* comprises papers presented at the second international congress of peer review held in Chicago in September 1993. A previous theme issue of *JAMA* published many of the papers from the first congress on peer review held in 1989.[9] Its aims were "to stimulate scientists to investigate those aspects of peer review that intrigued them, to discuss relevant and pressing issues in peer review, and to throw light on what has become one of the most important quality control mechanisms in science."[10]

About 70% of the 35 papers selected for presentation were the result of investigations; the remainder were opinion pieces. Introducing the papers in an editorial in the theme issue, Drummond Rennie wrote: "Though we certainly believe that we achieved our objectives, it is obvious that we have only begun to scratch the surface."[10]

The second congress attracted 110 papers; many more of them were based on investigation rather than opinion. A research base is thus beginning to emerge, but, as Jerry Kassirer, editor of the *New England Journal of Medicine*, observes in a paper in the second theme issue: "Of the articles published from the first peer review congress, all but one addressed manuscript management, not manuscript assessment."[11] This is also largely true of the second congress, and people at the congress observed that most of the research was descriptive and analytical rather than based on experimental intervention. It is as if treatment in medicine had to be based on analytical studies of existing practice rather than randomised clinical trials.

Great progress has been made with research into peer review, but we still have a long way to go. We can see this, for example, with studies of blinding peer reviewers. A paper presented at the first congress showed that blinding could produce a significant improvement in the quality of reviewers' opinions.[12] But this was a small study conducted in one journal, and the authors thought that it would be premature for editors to switch to blinded peer review on the basis of this single study. Two papers presented at the second congress also suggested that blinding reviewers produces benefits,[13,14] but again the evidence is still lacking to begin what may be a complicated process. It may also be that other processes may be more effective at raising the quality of referees' opinions—for instance, training them, removing their anonymity,[15] or providing them with guidelines or feedback. A protocol has been designed to conduct an international multijournal trial of blinding reviewers, although an alternative strategy has also been suggested of different journals conducting trials of different methods—only in a way that will allow the results to be compared in a systematic review.

In an attempt to promote research into a peer review, particularly in Europe, the editors of the *BMJ* and the *Lancet* and many others have formed a research network. The network currently has about 60 members, and our aims are to develop international research projects into the preparation, publication, and dissemination of health research, and to raise its quality; to produce a database of projects under way; to develop projects for presentation at the third international congress; to bring together people interested in this sort of research; to encourage an environment in which organisations are interested to fund such research and people are keen to engage in it; to promote the idea that such research is worthwhile and important; and to ensure that the results of the research are fed through into practice. After our first meeting we set up seven groups to look into different aspects of the preparation, publication, and dissemination of health research: how decisions are made on which papers to publish; measuring the outcomes of the processes; the problems of authors; the special issues surrounding specialist rather than general journals; the avoidance of fraud; the dissemination of results through the mass media; and the importance of the drug industry in the processes. We invite anybody interested in any of these activities to join us. You should contact me.

These activities may have a long term importance in that they are an important part of medical editing becoming more professional. Professions are characterised by developing a knowledge base, ensuring that that knowledge is used to improve practice, and developing a mechanism for spurring on those who don't reach high standards. The peer review congress, for which *JAMA* and Drummond Rennie must be thanked and admired, addressed all of these issues, including the last: three Britons argued the case for the creation of an international medical scientific press council.[16]

RICHARD SMITH
Editor

BMJ,
London WC1H 9JR

1 Lock S. *A difficult balance: editorial peer review in medicine.* London: The Nuffield Provincial Hospitals Trust, 1985. (Reprinted London: BMJ Publishing Group, 1991.)
2 Fletcher SW, Fletcher RH. Peer review and medical journals: how to make a good thing better? *National Medical Journal of India* 1994;**7**:103-5.
3 Smith R. Problems with peer review and alternatives. *BMJ* 1988;**296**:774-7.
4 Williamson JW, Goldschmidt PG, Colton T. The quality of medical literature: an analysis of validation assessments. In: Bailar JC, Mosteller F. *Medical uses of statistics*. Waltham, MA: NEJM Books, 1986.
5 Altman D. The scandal of medical research. *BMJ* 1994;**308**:283-4.
6 Rennie D, Flanagin A. The second international congress on peer review in biomedical publication. *JAMA* 1994;**272**:91.
7 Rennie D, Knoll E. Investigating peer review. *Ann Intern Med* 1988;**109**:181.
8 Bailar JC, Patterson K. Journal peer review: the need for a research agenda. *N Engl J Med* 1985;**312**:654-7.
9 Guarding the guardians: research on editorial peer review. Selected proceedings from the first international congress of peer review in biomedical publication, 10-12 May, Chicago, Ill. *JAMA* 1990;**263**:1317-414.
10 Rennie D. Guarding the guardians: a conference on editorial peer review. *JAMA* 1986;**256**:2391-2.
11 Kassirer JP, Campion EW. Peer review: crude and understudied, but indispensable. *JAMA* 1994;**272**:96-7.
12 McNutt RA, Evans T, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review. *JAMA* 1990;**263**:1371-6.
13 Fisher M. Friedman SB, Strauss B. The effects of blinding on acceptance of research papers by peer review. *JAMA* 1994;**272**:143-6.
14 Laband DN, Piette MJ. A citation analysis of the impact of blinded peer review. *JAMA* 1994;**272**:147-9.
15 Fabiato A. Anonymity of reviewers. *Cardiovasc Res* 1994;**28**:1134-9.
16 Altman DG, Chalmers I, Herxheimer A. Is there a case for an international medical scientific press council? *JAMA* 1994;**272**:166-7.