Assembling comparison groups to assess the effects of health care*

Iain Chalmers MSc

J R Soc Med 1997;90:379–386

Ask what was Bradford Hill's most important specific contribution to knowledge and the answer must surely be his research with Richard Doll on the relation between tobacco and lung cancer. But medical numerophobes such as myself honour him also for the clarity of his writing for non-statisticians. This is nowhere better exemplified than in his two articles entitled 'The Clinical Trial', published at the beginning of the 1950s[1,2]. In these papers, he notes the circumstances in which carefully controlled trials are unnecessary; he discusses the ethics of doing and of not doing trials; and he covers virtually all the methodological aspects of the subject matter that are judged important today. He even remarks, for those who perceive some inherent antithesis between controlled trials and the collection of qualitative data, that as long as the studies have been appropriately designed to control biases, subjective impressions can be given full weight in analyses of controlled trials. And he also has important things to say about the relevance of the results of controlled trials in practice:

> . . . [I]t appears sometimes to be thought that there is some necessary antagonism between the clinical assessment of a few cases and the "cold mathematics" of the statistically analysed trial dealing with a larger number. It is difficult to see how in fact there can be any such antagonism. The clinical assessment, or the clinical impression, must itself be numerical in the long run—that patients are reacting in a way different from the way the clinician believes was customary in the past. In the controlled trial an attempt is made to systematize those impressions (and other measurements) and to add them up. Standard errors, alarming as they may be to some persons, are wholly subsidiary, being merely tests, when the answer has been reached, as to whether it is safe to generalize from that answer[2].

This point is relevant to the often fruitless exchange of opinions about the extent to which the results of randomized trials are applicable in particular settings.

After Bradford Hill, it was one of his students Archie Cochrane who was outstandingly the most effective communicator of the rationale for properly controlled trials of health care interventions. In particular, Cochrane's book *Effectiveness and Efficiency: Random Reflections on Health Services*[3], published 25 years ago, had an impact far beyond the confines of the medical profession. The introductory chapter concludes by referring to the 1948 report of the randomized trial of streptomycin for pulmonary tuberculosis[4] and by paying tribute to Bradford Hill:

> It was an important paper in many ways, but from the point of view of the NHS it enabled Bradford Hill to introduce to the medical world the techniques of the RCT [randomized controlled trial] which added the experimental approach to medical research. Its importance cannot be exaggerated. It opened up a new world of evaluation and control which will, I think, be the key to a rational health service[3].

During the subsequent decade Cochrane promoted this view vigorously; and it was in an essay that he wrote for the Office of Health Economics, published in 1979, that he issued the challenge that has stimulated the emergence of the international Collaboration that bears his name:

> It is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials[5].

Cochrane's reference to the need for a 'critical summary' reflected two insufficiently acknowledged realities—first, that a synthesis of all the research evidence relevant to a particular question, and not the individual studies, constitutes the substantive output of the health research enterprise; and, second, that those preparing such syntheses must take steps to minimize biases and take account of the effects of the play of chance just as they are expected to do when collecting and analysing new data in 'primary' research.

Explicit demonstration and acknowledgement of the poor scientific quality of reviews of health research really began only in the late 1980s with the publications of Mulrow[6] and Oxman[7]. Relevant work had, however, been started some years earlier, particularly by Tom Chalmers, Richard Peto, and their respective colleagues, and by an international collaborative group preparing systematic reviews relevant to care during pregnancy and childbirth. Systematic reviews of a high proportion of all the relevant controlled trials in pregnancy and childbirth began to appear in journal articles, in books, and in an electronic journal in

UK Cochrane Centre, NHS Research & Development Programme, Summertown Pavilion, Middle Way, Oxford OX2 7LG, England

*Based on the 6th Bradford Hill Memorial Lecture, London School of Hygiene and Tropical Medicine, 30 April 1997

which the reviews were updated as new evidence became available and errors were identified[8].

In 1987, the year before he died, Cochrane referred to the collection of systematic reviews of controlled trials in pregnancy and childbirth as 'a real milestone in the history of randomized controlled trials and in the evaluation of care,' and he suggested that other specialties should copy the methods that had been used[9].

How might this challenge be met? An opportunity emerged with the inauguration of the Research and Development Programme established in 1991 to support the National Health Service[10]; and funds were provided to establish a Cochrane Centre to facilitate systematic reviews of randomized trials across all areas of health care[11,12]. When the centre opened in Oxford in October 1992, those involved expressed the hope that there would be a collaborative international response to Cochrane's agenda[12,13]; and, a year later, 77 people from nine countries gathered to co-found the Cochrane Collaboration. The Collaboration has evolved rapidly and many hundreds of people are now involved. Most contribute through about 40 collaborative review groups, each assessing interventions for prevention, treatment and rehabilitation for specific health problems. The work of these collaborative review groups is coordinated and supported by over a dozen Cochrane centres and a similar number of methodological working groups[14]. The enterprise is being built on eight values—collaboration, building on the enthusiasm of individuals, avoiding duplication, minimizing bias, keeping up to date, ensuring relevance, ensuring access, and continually improving the quality of its work[15].

In this paper I focus on just one aspect of one of these—ways to minimize biases that can arise during assembly of comparison groups to assess the effects of health care interventions.

ASSEMBLING COMPARISON GROUPS TO ASSESS THE EFFECTS OF HEALTH CARE

Although the Medical Research Council's randomized trial of streptomycin for pulmonary tuberculosis in 1948 is properly regarded as a landmark in medical history[4], it was by no means the beginning of the modern era of controlled trials. For example, Vandenbroucke[16] notes that the report of a trial on the serum treatment of lobar pneumonia, conducted under the auspices of a therapeutic trials committee of the Medical Research Council which had been established in 1931, contains 'a beautiful discussion of selection and comparability of treatment groups.' Here is part of it:

> The good results of insulin on patients with diabetes or of liver treatment in pernicious anaemia are so constant that the trial of these remedies in a very few cases was enough to establish their value. With the antiserum treatment of lobar pneumonia the conditions are very different. . . . If a straightforward comparison of treated cases with controls, under the average conditions whereby patients succeed one another in the wards of a hospital, could not reveal any advantage for those treated by serum, then common sense would conclude that the use of this remedy should be disregarded in the routine of practical medicine. The method consequently agreed upon . . . was that alternate cases of lobar pneumonia, taken simply in the order of their admission to hospital, should be used respectively for serum treatment and controls[17].

When Vandenbroucke's article was published in 1987, I sent a copy to Bradford Hill and he replied that he felt certain he had written the paragraph concerned. Bradford Hill's decision to use alternation rather than formal randomization in this and other trials during the 1930s was deliberate. He thought it would be easier for clinicians to conceptualize the way that strict alternation would result in 'a random division of the patients among the comparison groups in a trial', as long as 'no departure from this rule [was] allowed'[18], and reiterated this view the year before he died[19]. Armitage[20] suspects that another reason for Bradford Hill's failure initially to distinguish clearly between alternation and randomization may have been that he underestimated the danger of selection bias which can arise from departures from strict alternation.

In principle, there is no reason why alternation should not result in the assembly of unbiased comparison groups. It is important to be clear that there is nothing inherently superior in using random numbers rather than alternation as a basis for generating unbiased comparison groups in controlled trials. Just as departures from strict alternation can introduce bias, so also can departures from schedules based on a list of random numbers or coin tosses. The key issue is whether or not the schedule is known or predictable by those involved in allocation to the comparison groups.

Uncertainty about the relative merits of alternative forms of care is the fundamental ethical principle upon which controlled trials must be built. When that uncertainty is not present, and the allocation schedule is known or predictable, people may either be excluded from or preferentially placed in a particular comparison group, depending on judgments about their prognoses and their prospects of benefiting from or being harmed by one or other of the treatments being compared.

It was Tom Chalmers and his colleagues[21] who first showed that failure to conceal the treatment allocation schedule from those recruiting participants to controlled trials was associated with maldistribution of prognostic factors, and with estimates of treatment differences that were larger and more likely to be statistically significant than those derived from trials in which the allocation schedule seemed likely to have been concealed. In a more recent analysis, Schulz and his colleagues[22] showed that

trials in which the allocation schedule had not been concealed yielded odds ratios that were 41% larger than those derived from trials in which authors had reported adequately concealed allocation schedules.

Black[23] has suggested that these latter findings may in part reflect an association between the extent of allocation concealment and the strength of placebo effects. Placebo effects may be affected when health care professionals are required to openly acknowledge their uncertainties about the effects of health care when offering treatments in the context of a randomized trial, but not when offering exactly the same treatments in other contexts[24].

The results of the analyses reported by Chalmers and Schulz and their colleagues are certainly consistent with Armitage's suggestion that Bradford Hill may have underestimated the frequency with which investigators might depart from strict alternation in controlled trials. The temptation to subvert allocation schedules in controlled trials in this way may be understandable if the basic requirement of uncertainty has not been met[25]; but such subversion will inevitably increase the risk that trials will yield biased estimates of the relative merits of the forms of care being compared. Because existing evidence suggests that these biases result not only in useless or frankly dangerous forms of care being deemed useful, but also in useful forms of care being overlooked, controlling them presents ethical as well as methodological challenges.

The challenges implied by this evidence have been addressed by the Cochrane Collaboration in two ways. The first is reflected in the criteria it has adopted for identifying and registering studies which are potentially relevant to its objectives. A study is eligible for inclusion in *The Cochrane Controlled Trials Register* if, on the basis of the best available information, the individuals (or other units) followed in the trial were definitely or possibly assigned prospectively to one of two (or more) alternative forms of health care by random allocation or some quasi-random method of allocation (such as alternation, date of birth, or case record number)[26].

The first and by far the most challenging stage of data collection for any systematic review involves identifying as high a proportion as possible of all the potentially relevant studies. This task is made particularly difficult by the host of reporting biases that are known to affect the accessibility of reports of research. Compared with studies yielding unremarkable estimates of the differential effects of treatments, studies that have yielded dramatic estimates are more likely to be selected for presentation at scientific meetings[27]; more likely to be submitted for publication[28,29]; more likely to be accepted for publication[30]; more likely to be published as full reports[31]; more likely to be published in journals that are widely read[32,33]; more likely to be published in English[34]; and more likely to be cited in reports of subsequent, related studies[35].

In the pilot project for the Cochrane Collaboration, great efforts were made—over a period of ten years—to overcome reporting biases[36–38]. In addition to the usual searches of bibliographic databases, the contents of over sixty relevant journals were examined page by page, going back to the first issues or to those published in 1950, whichever was the earlier. Conference proceedings were also scrutinized. And letters were sent to over 40 000 people in an attempt to flush out information about unpublished studies. These efforts have been expanded and intensified with the emergence of the Cochrane Collaboration. Nearly a thousand journals have been or are being handsearched for reports of controlled trials; and the numbers of controlled trials recognizable as such on MEDLINE has quadrupled as a result of the Collaboration's work over the past three years[39]. As a result of all this work, *The Cochrane Controlled Trials Register* currently holds 130 000 records, many of which are not available in any other bibliographic database, and it has thus become the best single source of references to controlled trials.

The second way in which people within the Cochrane Collaboration are attempting to minimize the effects of bias in assembling comparison groups is by assessing the likelihood that bias has occurred in each of the studies being considered for possible inclusion in a review. In areas of research in which a large number of trials have been reported, this may mean restricting the review to trials in which the allocation schedule was concealed. In other areas, this restriction would mean that there were few or no trials to analyse. In these circumstances, as in the pilot work in pregnancy and childbirth[40,41], those preparing reviews use software enabling them to order the studies contributing to their review on the basis of their assessment of the likelihood of bias in each study. The same software is used to display the reviews in *The Cochrane Database of Systematic Reviews*.

RANDOMIZED AND NON-RANDOMIZED COMPARISONS FOR ASSESSING THE EFFECTS OF HEALTH CARE

Some people have criticized the Cochrane Collaboration for its implicit encouragement of those preparing Cochrane reviews to consider studies in which comparison groups have been assembled by methods that are open to bias—such as alternation. Far greater numbers, however, have been critical of the Collaboration for what they perceive to be an inappropriate obsession with the importance of randomization in controlling bias[23].

At one level, this suggestion is consistent with the explicitly stated aims of the Cochrane Collaboration. As stated at its inception in 1993, the Collaboration has been established 'to prepare, maintain and disseminate systematic, up-to-date reviews of randomized controlled trials

(RCTs) of health care, and, when RCTs are not available, reviews of the most reliable evidence from other sources'. One Cochrane collaborative review group, for example, is registering and considering studies using interrupted time series and controlled before-and-after designs[42], and some Cochrane reviews have incorporated evidence from non-randomized studies to assess hypotheses about rare adverse effects of treatment[43].

Calls on the Cochrane Collaboration to include systematic reviews of non-randomized comparisons of alternative forms of care are uncontentious in circumstances in which it is inconceivable that randomized trials will ever provide the information that members of the public require. For example, it is unlikely that randomized cohorts of women allocated to hormonal or barrier methods of contraception could ever be sustained for long enough to assess the long-term effects of oral contraceptives on cardiovascular disease and cancer.

The main differences of opinion seem to be about whether or not to use non-randomized comparison groups in circumstances in which randomized trials *can* be done, but where they have either not been done or have not been done to a sufficiently high standard. As two social scientists, Ann Oakley[44] and Geraldine Macdonald[45], have pointed out particularly eloquently, the reasons sometimes given for rejecting randomized trials are disingenuous. But rejection of randomized trials may often simply reflect ignorance of the range of circumstances in which randomization has been and therefore plainly can be used to minimize bias; or insufficient acknowledgment of the ways that bias can lead to the adoption of ineffective or dangerous forms of care; or lack of appreciation of the modest prior probability of a proposed new form of health care turning out to be superior to existing options[46].

Until very recently, for example, the drug of choice for controlling eclamptic convulsions was unclear because there were no controlled trials. Arguments had gone back and forth for nearly a century because of the impossibility of interpreting uncontrolled case series and non-randomized comparisons of different drugs. In 1995, publication of the results of the work of the Eclampsia Trial Collaborative Group resolved these uncertainties[47], but not before literally millions of women had paid the price of the failure to begin doing randomized trials decades ago.

Sometimes reports of randomized trials are available but the studies have not been done to a sufficiently high standard to provide a basis for guiding choices in health care. For example, there are well over 500 potentially relevant controlled trials of a wide variety of drugs used to prevent or reduce the unpleasant movement disorders suffered by people taking anti-psychotic drugs. Collectively, this vast body of research yields *no* information upon which practice can be based with any confidence[48]. The research is characterized by inadequate control of biases, inadequate sample sizes, inadequate length of follow-up, and the use of outcome measures which are irrelevant to patients and their families and others caring for them. It is one of many possible examples of what Altman[49] has referred to as the scandal of poor medical research.

When there are no informative randomized trials addressing important questions that could, in principle, be addressed in such trials, researchers and research funders are faced with a choice. Either they can invest resources in doing well-designed randomized trials, as they did to find out which anticonvulsant should be used in eclampsia; or they can invest resources in preparing systematic reviews of data from non-randomized comparisons, as has been done, for example, in assessing the relative merits of transurethral and open prostatectomy. If they judge the latter approach to be more cost-effective, the main difficulty they must confront is that the methodology for systematic reviews of data derived from non-randomized comparisons remains relatively undeveloped.

Whereas there is now a considerable and rapidly expanding body of empirical research to guide people preparing systematic reviews of randomized trials, there is far less guidance available to those who wish to prepare systematic reviews of non-randomized comparisons. Uncertainties surround which selection criteria should be used for such studies, how a high proportion of the studies that meet the selection criteria should be identified, and how to assess the comparability of the comparison groups, particularly when important prognostic factors may not have been identified or measured. It was because these issues could not be addressed satisfactorily that one recent attempt to make methodological progress in this field failed[50]. A further international meeting convened by the Centers for Disease Control and Prevention may result in progress, but some[5] have already endorsed Feinstein's view that the meta-analysis of non-randomized observational studies resembles the attempt of a quadriplegic person to climb Mount Everest unaided[52].

This lack of methodological groundwork is the main reason that some of those within the Cochrane Collaboration, including me, have been reluctant to encourage people to invest resources in trying to prepare systematic reviews of non-randomized comparisons in circumstances in which randomized trials could, in principle, be organized. Having said that, I should make clear that the methodological challenges are not being brushed aside within the Collaboration, if only because there is a clear consensus that methods must be developed to test hypotheses about rare adverse (or beneficial) effects of interventions, particularly those that have been shown in randomized trials to have useful effects. People interested in developing methodology in this area will be meeting at the forthcoming

Cochrane Colloquium in Amsterdam to discuss what empirical research might be done to inform future decisions.

THE POTENTIAL COSTS OF GETTING WRONG ANSWERS

When the results of randomized trials are compared with the results of non-randomized studies addressing the same questions the randomized trials tend to yield less striking differences (Kunz RA, Oxman A, unpublished; Britton A, McPherson K, McKee M, *et al*, unpublished)—like the results of trials in which allocation has been adequately concealed compared with those in which it has not[21,22]. How should researchers, health professionals and those responsible for taking funding decisions about health services and health research act when confronted with these differences, or when there are no randomized trials at all?

Sometimes a great deal is at stake. In 1981, for example, Horwitz and Feinstein[53] published a study entitled 'Improved observational method for studying therapeutic efficacy'. Their case–control analysis suggested that prophylaxis with lignocaine prevented both abnormalities of heart rhythm and death after heart attack. Lignocaine had been used for some time in clinical practice at the time the Horwitz and Feinstein study was published, and several other antiarrhythmic drugs became widely adopted in clinical practice during the 1980s. Two years after their paper had helped to promote wider use of these drugs, Furberg presented a systematic review of the randomized trials of prophylactic antiarrhythmic drugs in myocardial infarction. His analysis provided no support for the notion that these drugs reduced the risk of death, and indicated that they might actually increase it[54]. Furberg's analysis was dismissed by some of those who have been influential in promoting this class of drugs, but a few years later, Hine and his colleagues[55] published a further systematic review of randomized trials which showed a statistically significant increased mortality associated with lignocaine prophylaxis. Subsequent reports have shown that prophylactic use of this class of drugs increases the risk of death[56,57].

As Mervyn Susser, a student of Bradford Hill, has noted 'Our many errors show that the practice of causal inference . . . remains an art. Although to assist us we have acquired analytical techniques, statistical methods and conventions, and logical criteria, ultimately the conclusions we reach are a matter of judgement'[58]. Clearly, however, the consequences for the public's health and safety are substantial, whichever of the two sources of evidence—randomized or non-randomized—one judges preferable. Consider the evidence relating to antiarrhythmic drugs. Using the evidence derived from randomized trials, Moore has estimated that, at the peak of their use in the late 1980s, antiarrhythmic drugs were causing the sudden deaths after myocardial infarction of between 20 000 and 70 000 people every year in the USA alone[54]. By contrast, using the estimates of a beneficial effect on mortality derived from the Horwitz and Feinstein analyses of non-randomized comparisons, one would have to conclude that many otherwise preventable deaths had occurred because the antiarrhythmic drugs had not been adopted sufficiently widely.

Bradford Hill ended his 1952 article 'The Clinical Trial' by noting that 'randomised trials are not the only means of investigation and experiment, nor invariably the best way of advancing knowledge of therapeutics'[2]. Although he did not elaborate on this statement in that article, I feel certain that he must have been referring to circumstances in which treatment effects are large, other circumstances where randomized trials are inconceivable, and still other circumstances in which non-randomized studies are likely to be the only way of detecting very rare treatment effects.

It would be interesting to know, for example, how Bradford Hill would have reacted to the current worldwide promotion by the World Health Organization and the World Bank of directly observed treatment shortcourse (DOTS) anti-tuberculosis therapy. On 19 March this year, the World Health Organization issued a press release containing the following text:

> 'DOTS is the biggest health breakthrough of this decade, in terms of the lives we will be able to save' said Dr Hiroshi Nakajima, Director-General of the World Health Organisation. 'We anticipate that at least 10 million deaths from tuberculosis will be prevented in the next ten years with the introduction and extensive use of the DOTS strategy'.

What is the basis for WHO's and the World Bank's confidence in the effectiveness of this particular strategy for improving adherence to anti-tuberculosis therapy? Whatever else it may be, it is not based on the results of a large body of publicly available evidence derived from randomized trials. A search of *The Cochrane Library* for material relevant to 'directly-observed therapy' yields no hits among 130 000 references in *The Cochrane Controlled Trials Register*, and just one hit in *The Cochrane Database of Systematic Reviews*. This is a systematic review of strategies for promoting adherence to anti-tuberculosis treatment prepared by J Volmink (director of the South African Cochrane Centre) and P Garner (coordinating editor of the Cochrane Infectious Diseases Group). It yields the following results:

> Strategies found to be of benefit were reminder cards sent to defaulters, assistance of patients by lay health workers, monetary incentives offered to patients, and increased supervision of TB clinic

staff. It is not possible to determine from current trials whether health education by itself leads to better adherence to treatment. Even though directly-observed therapy (DOT) is widely advocated as the most cost-effective means of ensuring completion of TB treatment, no completed trials could be found which confirm or refute this view[59].

No-one can say, with confidence, that WHO and the World Bank are wrong to have launched their campaign without strong evidence from randomized trials that DOTS is likely to be the most effective and cost-effective strategy for controlling tuberculosis. But surely these influential organizations have a duty to make clearer the evidence base upon which their policies are being promulgated, if only because it is conceivable that they may do more harm than good.

SOME ETHICAL ISSUES

Writing as a patient, I have made clear in print that I wish the health professionals from whom I seek help to take into account, whenever possible, evidence assembled in systematic reviews of randomized trials; and, when these have revealed that uncertainty is justified, I want to be invited to participate in such trials[60]. I have also made clear that I believe that my interests and the interests of other patients and potential patients will be served more effectively when lay people become more involved in the design of such research[61,62].

It is consistent with what I want for myself that I believe the Cochrane Collaboration should concentrate its limited resources on preparing and maintaining systematic reviews of the hundreds of thousands of studies that have used the design that Bradford Hill played such an important role in introducing and promulgating. When there is insufficient evidence from randomized trials in circumstances where they are possible, I think that investment in randomized trials will often represent a more cost-effective use of scarce research resources than organizing and reviewing non-randomized comparisons. As Tom Chalmers put it, 'I don't think anybody in his right mind thinks for one minute that you can learn how patients *should* be treated by observing how doctors *are* treating them'[63].

It is certain that others, like Horwitz and Feinstein[53], will continue to explore how unbiased estimates of the effects of health care might be obtained using data from non-randomized comparisons in circumstances in which randomized trials are feasible. In my view, this diverts attention from some of the far more pressing challenges facing those seeking unbiased comparisons of alternative forms of health care.

I want to conclude by singling out just one among these challenges. In designing and interpreting the results of randomized trials, we need to understand better how to measure and take account of psychologically mediated effects of health care[24]. In particular, in interpreting randomized trials, we need to understand how the comparisons may be influenced by the fact that people in randomized trials receive different information about their treatment, communicated in different ways, from that received by people receiving exactly the same treatments outside the context of randomized trials. Testing hypotheses about the effects of this externally imposed double standard on informed consent to treatment[64] undoubtedly presents a methodological challenge. Perhaps it is even more important, however, to expose the double standard and challenge of those who promote it to justify their position, particularly as some of them now propose that journals should refuse to publish reports of randomized trials in which informed consent to treatment has not been sought[65].

In Britain, it was Maurice Pappworth[66] who first claimed moral high ground on this issue. Twice I challenged Pappworth in journal correspondence columns to state which of two groups of obstetricians he judged to have behaved more ethically[67,68]. A tiny minority of obstetricians, without seeking informed consent to randomize, conducted controlled trials of diethylstilboestrol in their pregnant patients. By so doing, they ensured that half would avoid the disastrous side-effects of the drug, and they gave up using it when no evidence of benefit had emerged from controlled trials by the mid-1950s. By contrast, thousands of obstetricians involved millions of pregnant women in the poorly controlled experimentation of what was 'accepted clinical practice' for over two decades. Accepted clinical practice, as is so often the case, was based on information about the efficacy and safety of the drug derived from non-randomized comparisons. Half a century later, the adverse consequences of this practice for women and their children are still reverberating. Pappworth never did respond to my challenge.

Some people seem to believe that randomization itself has some ethical connotations. Late in his life, Bradford Hill was prompted by a *BMJ* leading article to comment on this notion:

> It is perhaps widely believed . . . that randomization per se contributes to the ethical problems inherent in a clinical trial. In my view it plays no such part. A treatment is put forward for care or prevention of an illness—for example, large doses of vitamin C for the common cold. The clinician thinks that there is no evidence of its value (or hazards) and that the only way in which he can satisfy himself is to conduct a trial in which he will have two groups of patients, as similar as he can make them, one to be given the proposed treatment and the other to be given the accepted treatment of the day or a placebo. Presuming, of course, that he has obtained their informed consent, is he entitled to use his patients in that way? That is the ethical problem that he has to face and on which he must

make his decision. Randomisation plays no part in it. It comes in after his decision merely as a technique for producing the two groups in an unbiased way and securing as far as possible their comparability. It is no more than a technique, as are the means he will use in assessing the results of the trial—for example, using the stethoscope to listen to breath sounds in the bronchus.

In a lecture given in memory of his former colleague Marc Daniels, Bradford Hill commented on the draft code of ethics which had been published by the World Medical Association in 1963. He had this to say about informed consent to treatment:

Personally, and speaking as a patient, I have no doubt whatsoever that there are circumstances in which the patient's consent to treatment to taking part in a controlled trial should be sought. I have equally no doubt that there are circumstances in which it need not—and even should not—be sought. My quarrel is again with a code that takes no heed—and in dealing with generalities can take no heed—of the enormously varying circumstances of clinical medicine[69].

Circumstances and people do vary. And for some people, even if they may constitute only a minority, confident professional certainty is a more effective form of health care than the explicit admission of professional uncertainty that is required of those who are providing care within the context of randomized trials. On ethics as in so many other ways, Bradford Hill's ideas provide an enduring guide for researchers and professionals trying to ensure that health care does more good than harm.

Acknowledgments I am grateful to Doug Altman, Peter Armitage, Nick Black, Jan Chalmers, Mike Clarke, Lelia Duley, Paul Garner, Peter Gøtzsche, Jos Kleijnen, Bruce Kupelnick, Klim McPherson, Cath Milwain, Ann Oakley, Sandy Oliver, Andy Oxman, Nick Partridge, Maggie Pearson, Dave Sackett, Ken Schulz, Chris Silagy, Jimmy Volmink and Charles Warlow for comments on a draft of this paper.

REFERENCES

1 Hill, A Bradford. The clinical trial. *Br Med Bull* 1951;**7**:278–82

2 Hill, A Bradford. The clinical trial. *N Engl J Med* 1952;**247**:113–19

3 Cochrane AL. *Effectiveness and Efficiency: Random Reflections on Health Services*. London: Nuffield Provincial Hospitals Trust, 1972

4 Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *BMJ* 1948;**ii**:769–82

5 Cochrane AL. 1931–1971: a critical review, with particular reference to the medical profession. In: *Medicines for the Year 2000*. London: Office of Health Economics, 1979:1–11

6 Mulrow CD. The medical review article: state of the science. *An Intern Med* 1987;**106**:485–8

7 Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988;**138**:697–703

8 Chalmers I. Improving the quality and dissemination of reviews of clinical research. In: Lock S, ed. *The Future of Medical Journals: In Commemoration of 150 Years of the British Medical Journal*. London: British Medical Journal, 1991:127–46

9 Cochrane AL. Foreword. In: Chalmers I, Enkin M, Keirse MJNC, eds. *Effective Care in Pregnancy and Childbirth*. Oxford: Oxford University Press, 1989

10 Peckham M. Research and development for the National Health Service. *Lancet* 1991;**338**:367–71

11 Editorial. Cochrane's legacy. *Lancet* 1992;**340**:1131–2

12 Chalmers I, Dickersin K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. *BMJ* 1992;**305**:786–8

13 Chalmers I. The Cochrane Collaboration: preparing, maintaining and disseminating systematic reviews of the effects of health care. In: Warren KS, Mosteller F, eds. Doing more good than harm: the evaluation of health care interventions. *Ann NY Acad Sci* 1993;**703**:156–63

14 Stewart LA, Clarke MJ, on behalf of the Cochrane Working Group on Meta-analysis using Individual Patient Data. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Stat Med* 1995;**14**:2057–79

15 Chalmers I, Sackett D, Silagy C. The Cochrane Collaboration. In: Maynard A, Chalmers I, eds. *Non-random Reflections on Health Services Research: on the 25th Anniversary of Archie Cochrane's Effectiveness and Efficiency*. London: BMJ Books, 1997:231–49

16 Vandenbroucke JP. A short note on the history of the randomized controlled trial. *J Chron Dis* 1987;**40**:985–7

17 Therapeutic Trials Committee of the Medical Research Council. The serum treatment of lobar pneumonia. *BMJ* 1934;**1**:241–5

18 Hill, A Bradford. Principles of medical statistics XV. *Lancet* 1937;**1**:883–5

19 Hill, A Bradford. Memories of the British Streptomycin Trial in Tuberculosis. *Cont Clin Trials* 1990;**11**:77–9

20 Armitage P. Bradford Hill and the randomized controlled trial. *Pharm Med* 1992;**6**:23–37

21 Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. *N Eng J Med* 1983;**309**:1358–61

22 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408–12

23 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;**312**:1215–18

24 Kleijnen J, Gotzsche P, Kunz RH, Oxman AD, Chalmers I. So what's so special about randomisation? In: Maynard A, Chalmers I, eds. *Non-random Reflections on Health Services Research: on the 25th Anniversary of Archie Cochrane's Effectiveness and Efficiency*. London: BMJ Books, 1997:93–106

25 Schulz K. Subverting randomization in controlled rials. *JAMA* 1995;**274**:1456–8

26 Dickersin K. Establishing and maintaining an international register of RCTs. In: Sackett D, Oxman A, eds. *Cochrane Collaboration Handbook*, Oxford: Cochrane Collaboration, 1994

27 Koren G, Graham K, Shear H, Einarson T. Bias against the null hypothesis: the reproductive hazards of cocaine. *Lancet* 1989;**ii**:1440–2

28 Dickersin K, Min YI. NIH clinical trials and publication bias [article]. *Online J Curr Clin Trials* [serial online] Apr 28; 1993 (Doc No 50): [4967 words; 53 paragraphs]. 1 fig; 3 tables

29 Simes RJ, Stern J, Ghersi D. The importance of prospectively registering trials: evidence of publication bias from a cohort study of research protocols. *BMJ* (in press)

30 Mahoney MJ. Bias, controversy, and abuse in the study of the scientific publication system. *Sci Tech Hum Values* 1990;**15**:50

31 Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. *JAMA* 1994;**272**:158–62

32 Simes J. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;**4**:1529–41

33 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;**337**:867–72

34 Egger M. Language reporting bias among German-speaking authors of reports of controlled trials. *Lancet* (in press)

35 Gøtzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;**295**:654–6

36 Chalmers I, Hetherington J, Newdick M, Mutch L, Grant A, Enkin M, Enkin E, Dickersin K. The Oxford Database of Perinatal Trials: developing a register of published reports of controlled trials. *Controlled Clin Trials* 1986;**7**:306–24

37 Chalmers I. Evaluating the effects of care during pregnancy and childbirth. In: Chalmers I, Enkin M, Keirse MJNC, eds. *Effective Care in Pregnancy and Childbirth*. Oxford: Oxford University Press, 1989: 3–38

38 Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989;**84**:374–80

39 McDonald SJ, Lefebvre C, Clarke MJ. Identifying reports of controlled trials in the *BMJ* and the *Lancet*. *BMJ* 1996;**313**:1116–17

40 Chalmers I, Hetherington J, Elbourne D, Keirse MJNC, Enkin M. Materials and methods used in synthesizing evidence to evaluate the effects of care during pregnancy and childbirth. In: Chalmers I, Enkin M, Keirse MJNC, eds. *Effective Care in Pregnancy and Childbirth*. Oxford: Oxford University Press, 1989:39–65

41 Chalmers I. On applying overviews and meta-analyses at the bedside: comments on David Sackett's paper. *J Clin Epidemiol* 1995;**48**: 67–70

42 Cochrane Collaboration on Effective Professional Practice. Module in *The Cochrane Library* [database on disk and CD-ROM]. The Cochrane Collaboration; Issue 2. Oxford: Update Software; 1997. Updated quarterly

43 Olliaro P, Mussano P. Amodiaquine for the treatment of malaria. In: Garner P, Gelband H, Olliaro P, Salinas R, Volmink J, Wilkinson D. Infectious Diseases Module of *The Cochrane Database of Systematic Reviews*. Available in *The Cochrane Library* [database on disk and CD-ROM]. The Cochrane Collaboration; Issue 2. Oxford: Update Software; 1997. Updated quarterly

44 Oakley A, Fullerton D. The lamp-post of research: support or illumination? The case for and against randomised controlled trials. In: Oakley A, Roberts H, eds *Evaluating Social Interventions*. Ilford: Barnardos, 1996:4–38

45 Macdonald G. Social work: beyond control? In: Maynard A, Chalmers I, eds. *Non-random Reflections on Health Services Research: on the 25th Anniversary of Archie Cochrane's Effectiveness and Efficiency*. London: BMJ Books, 1997:122–46

46 Chalmers I. What is the prior probability of a proposed new treatment being superior to established treatments? *BMJ* 1997;**314**:74–5

47 Eclampsia Trial Collaborative Group. Which anticonvulsant for women with eclampsia? *Lancet* 1995;**345**:1455–63

48 Soares K, McGrath J, Adams C. Evidence and tardive dyskinesia. *Lancet* 1996;**347**:1696–7

49 Altman DG. The scandal of poor medical research. *BMJ* 1994;**308**:283–4

50 Spitzer WO, ed. The Potsdam International Consultation on Meta-analysis. *J Clin Epidemiol* 1995;**48**:1–172

51 Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *J Clin Epidemiol* 1997;**50**:223–9

52 Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 1995;**48**:71–9

53 Horwitz RI, Feinstein AR. Improved observational method for studying therapeutic efficacy: suggestive evidence that lidocaine prophylaxis prevents death in acute myocardial infarction. *JAMA* 1981;**246**:2455–59

54 Moore T. *Deadly Medicine*. New York: Simon and Schuster, 1995.

55 Hine LK, Laird N, Hewitt P, Chalmers TC. Meta-analytic evidence against prophylactic use of lidocaine in acute myocardial infarction. *Arch Intern Med* 1989;**149**:2694–8

56 Cardiac Arrhythmia Suppression Trial Investigators. Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989;**321**:406–12

57 Cardiac Arrhythmia Suppression Trial II Investigators. Effect of antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med* 1992;**327**:227–33

58 Susser M. Causal thinking in practice: strengths and weaknesses of the clinical vantage point. *Pediatrics* 1984;**74**:842–9

59 Volmink J, Garner P. Promoting adherence to tuberculosis treatment. In: Garner P, Gelband H, Olliaro P, Salinas R, Volmink J, Wilkinson D. Infectious Diseases Module of *The Cochrane Database of Systematic Reviews*. Available in *The Cochrane Library* [database on disk and CD-ROM]. The Cochrane Collaboration; Issue 2. Oxford: Update Software; 1997. Updated quarterly

60 Chalmers I. What do I want from health research and researchers when I am a patient? *BMJ* 1995;**310**:1315–18

61 Hart JT. Response rates in South Wales 1950–96:changing requirement for mass participation in human research. In: Maynard A, Chalmers I, eds. *Non-random Reflections on Health Services Research: on the 25th Anniversary of Archie Cochrane's Effectiveness and Efficiency*. London: BMJ Books, 1997:31–57

62 Oliver S. Exploring lay perspectives on questions of effectiveness. In: Maynard A, Chalmers I, eds. *Non-random Reflections on Health Services Research: on the 25th Anniversary of Archie Cochrane's Effectiveness and Efficiency*. London: BMJ Books, 1997:272–91

63 Chalmers TC. In: Warren KS, Mosteller F, eds. Doing more good than harm: the evaluation of health care interventions. *Ann NY Acad Sci* 1993;**703**:122

64 Chalmers I, Silverman WA. Professional and public double standards on clinical experimentation. *Controlled Clin Trials* 1987;**388**:91

65 Doyal L. Journals should not publish research to which patients have not given fully informed consent—with three exceptions. *BMJ* 1997;**314**:1107–11

66 Pappworth MH. Human guinea pigs: a history. *BMJ* 1990;**301**:1456–60

67 Chalmers I. Medical experimentation. *World Med* 1978 April 5:18

68 Chalmers I. Human guinea pigs. *BMJ* 1991;**302**:411

69 Hill, A Bradford. Medical ethics and controlled trials. *BMJ* 1963;**1**:1043–9