Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

2

The lethal consequences of failing to make full use of all relevant evidence about the effects of medical treatments: the importance of systematic reviews

Iain Chalmers

Introduction

For nearly two centuries there have been arguments about the relevance of evidence derived from research involving groups of patients when treatment decisions are needed for individual patients.[1, 2] This book will help to promote much-needed clearer thinking about this issue. Only rarely is complete certainty justified about how a treatment will affect a particular patient. Half a century ago, not long after the birth of the randomised clinical trial, Austin Bradford Hill[3] drew attention to the inevitable guesswork involved in using the results of clinical trials to predict the effects of treatments in individuals:

> *Our answers from the clinical trial present ... a group reaction. They show that one group fared better than another, that given a certain treatment, patients, on the average, get better more frequently or rapidly, or both. We cannot necessarily, perhaps very rarely, pass from that to stating exactly what effect the treatment will have on a particular patient.* But there is, surely, no way and no method of deciding that. [My emphasis.][3]

How do I want this inevitable guesswork about the effects of treatment to be informed when I am a patient?[4] Very occasionally it will be possible to design research ('*n*-of-1' trials) to find out which of alternative treatments best suits me – absolutely specifically. More usually this kind of very specific information will not be available, even when my genotype becomes known.[5] Patients like me, and the clinicians to whom we look for personalised care, are clearly very interested in knowing which factors should be taken into account when assessing the applicability of average effects of treatments derived from studying groups of patients. Although it seems unlikely that 'just about every treatment does some good for someone', as some have suggested,[6, 7] I imagine I am not alone in wanting decisions about my treatment to be as individualised as possible. This is a tough challenge and later chapters in this book will deal with twin thorny problems: identifying which individuals are likely to benefit from or be harmed by treatments, and avoiding the false inferences that can result from biases and chance associations.[8–10]

The widely promulgated vision of general availability of individually tailored treatments seems still to be some way from being realised in practice. During the lifetime of this

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

edition of this book at least, my clinician advisers and I will usually need to fall back on evidence derived from observing groups of patients who are more or less like me. Often this evidence will be informal, existing only inside the heads of my advisers, who will draw on their experience of treating other patients. I may want that kind of informal evidence to be taken into account in decisions about my treatment; but as there are many examples of such evidence having been dangerously misleading, I certainly want account to be taken of relevant research evidence.

Some people may suggest that I should not assume that evidence from research is relevant to me unless the people studied in the research can be shown to be very like me. My approach is different. I want to know whether there are reasons that could justify confidently dismissing – as irrelevant to me – estimates of effects derived from the best available research evidence on groups of people. These estimates might be dismissed either because I am confident that I am completely different from the people who participated in the research; or because the interventions available to me are not those that have been evaluated by researchers; or because the questions or the treatment outcomes which I rate as important have been ignored by researchers.[11]

When I say that I want 'relevant research evidence' to be taken into account in decisions about my treatment, what kind of evidence do I have in mind? Although very large randomised trials may often contribute the overwhelming weight of evidence on particular therapeutic questions, this is no reason to ignore evidence from smaller studies judged likely to be unbiased, particularly if they have been registered prospectively to reduce publication bias.[12] I want treatment decisions to be informed by synthesis of all the relevant evidence – synthesised rigorously in systematic reviews.

And I do mean *all* the relevant evidence. Biased underreporting of research can be lethal. For example, Cowley et al.,[13] to their great credit, reported in 1993 the results of a clinical trial of an antiarrhythmic drug in myocardial infarction actually done 13 years earlier.

> *Nine patients died in the lorcainide (drug) group and one in the placebo group ... When we carried out our study in 1980 we thought that the increased death rate that occurred in the lorcainide group was an effect of chance ... The development of lorcainide was abandoned for commercial reasons, and this study was therefore never published; it is now a good example of 'publication bias'. The results described here ... might have provided an early warning of trouble ahead.*[13]

The 'trouble ahead' to which they were referring was that, at the peak of their use, this class of drugs was causing tens of thousands of premature deaths every year in the USA alone.[14] Worldwide, the drugs seem likely to have caused hundreds of thousands of deaths.

Although most of the evidence of the dangers of biased underreporting of research has been sought and found among practice-orientated clinical research studies,[15] it appears

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

that the results of an unpublished study might have provided a forewarning of the tragic consequences of using the drug TGN1412, which had life-threatening effects on six young, healthy volunteers in study in London in early 2006.[16] Biased underreporting of preclinical research may also help to explain the high and increasing failure of proposals for new drugs to survive assessment in early clinical studies.[17]

As far as possible, therefore, whether in early or late studies of treatments, systematic reviews should take account of *all* the relevant data – published and unpublished. Systematic reviews based on these data can show that when the results of studies addressing the same or similar questions appear to differ, the apparent differences (even in the results of large studies) are compatible with the effects of chance.[18] And when chance is an unlikely explanation for differing results of apparently similar trials, systematic reviews can be used to explore and possibly explain the differences in ways that may improve understanding of how to individualise treatment. Systematic reviews can also be used to test hypotheses generated by unexpected results in individual trials. For example, after a trial had found an unpredicted, statistically significantly higher rate of breast cancer in women who had received a lipid-lowering drug,[19] a systematic review of all similar studies provided reassurance that the observation was very likely to have reflected chance.[20]

Of course, the term 'systematic review' begs the question – what system? As with reports of any scientific investigation, the expectation of readers should be that reports of systematic reviews will contain descriptions of the materials and methods used by researchers to address clearly stated questions, in ways that reduce distortions from biases and the play of chance. The term 'systematic review' means for me 'the application of strategies that *limit bias* [my emphasis] in the assembly, critical appraisal, and synthesis of all relevant studies on a specific topic'.[21] Meta-analysis – 'statistical synthesis of the data from separate but similar studies leading to a quantitative summary of the pooled results'[21] – may or may not be a component of systematic reviews, but it does nothing to reduce bias. If appropriate and possible, however, it can often reduce the likelihood of our being misled by the effects of the play of chance, as Karl Pearson demonstrated more than a century ago.[22]

For questions about the effects of healthcare interventions, a key issue concerns the kind of primary studies that will be eligible for inclusion in systematic reviews. Some interventions have dramatic effects that can be confidently identified without carefully controlled research.[23] Unbiased, confident detection of the more modest effects of most interventions, however, requires sufficiently large studies that have used randomisation and other measures to minimise biases. That is why this book emphasises the importance of randomised trials. There are many examples of the dangers of basing treatment decisions on non-randomised studies. An example relevant to a later section in this chapter is a non-randomised study reported by Horwitz and Feinstein:[24] the results of their analysis encouraged the use of a class of drugs that turned out to be lethal.[14]

Because people have not taken sufficiently seriously the need to make full use of *all* the relevant research evidence, readers of research reports have been misled by biases and the play of chance, patients have suffered and died unnecessarily, and resources for

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

healthcare and health research have been wasted. It was against this background that Peter Rothwell asked me to stress in this chapter 'the absolute importance of research synthesis in determining the overall effects of treatment, and the fact that this has to be the starting point for any further consideration of who might benefit most.' (P. Rothwell, personal communication, 2 September 2005).

The scandalous failure of biomedical science to cumulate evidence scientifically

Four decades ago, one of the pioneers of randomised trials, Austin Bradford Hill, suggested that readers of published reports of research want their authors to provide the answers to four basic questions: 'Why did you start; what did you do; what did you find; and what does it mean anyway?'[25] As a result of the adoption of reporting guidelines such as those recommended by the CONSORT Group,[26] it has become more likely that readers of research reports will have satisfactory answers to Hill's second and third questions – What did you do? and What did you find? – but satisfactory answers to Hill's first and fourth questions are much rarer.

To answer Bradford Hill's first question – Why did you start? – readers need to be reassured that new research has been done because, at the time it was initiated, important questions could not be answered satisfactorily with existing evidence. Systematic reviews of existing evidence have sometimes raised serious ethical questions, for example, about the continued use of placebos or no treatment controls.[27, 29] Although these problems have been exposed repeatedly over the past quarter century, a recent survey showed that researchers still do not, in general, review existing evidence systematically when designing controlled trials (30). Indeed, some of them state bluntly that they see no need to do this (A. Sutton, personal communication, **{AQ_01}**). Even among those who take a formal Bayesian approach to the design of controlled trials, the fairly basic of step of using a systematic review of existing evidence to estimate the likely size of a treatment difference seems often to be overlooked.[31] Although applicants to some research funding bodies are required[32] or being urged[33] to show how their proposals build on systematic reviews of existing evidence, the unjustified duplication of research illustrated in the examples given later in this chapter make clear that many research funders are not taking sufficiently seriously their responsibilities to husband limited resources for research efficiently and ethically.

Research funding organisations, research ethics committees, journals and drug licensing authorities have all been challenged to accept the responsibilities implied by their authority to reduce this form of research misconduct.[34, 35] The situation would improve if there was less institutionalised reverence for the current systems of peer review operated by research funders and editors of journals. The failure of authors to mention relevant previous work is a widespread form of research malpractice.[36] Fifteen years ago, Chalmers et al.[37] suggested that, when submitting reports for review, investigators should be required by journals to provide evidence that they have made a thorough search for relevant previous work. Such a requirement might both improve the relevance and quality of research conducted and help to reduce the frequency of undeclared duplicate publication, as well as improve detection of plagiarism.[38] Yet very few funders and journal

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

editors appear to have taken seriously that suggestion, or Jefferson and co-workers'[39, 40] related proposal that, until proved otherwise, a research proposal or a submitted manuscript should be viewed as one member of a population of related studies, and judged in the context of a systematic review of the other member studies in that population.

Research ethics committees have done little to protect patients from the adverse effects of this scientific sloppiness. Ten years have passed since research ethics committees were challenged publicly to recognise that they were behaving unethically if they did not take steps to satisfy themselves that they were approving research that had been designed in the light of systematic reviews of existing evidence;[34] yet there is very little evidence that they have taken this challenge seriously.[41, 42] The failure of research ethics committees to hold research funders and researchers to account in this respect can sometimes result in dramatic tragedies. For example, Ellen Roche, a young woman volunteer in a physiological experiment at Johns Hopkins Medical School, died because the design of the experiment in which she had been asked to participate had not been informed by a systematic review of relevant pre-existing evidence about the hazards of inhaled hexamethonium.[43] The researchers had depended on Medline in their search for relevant evidence about the effects of inhaling the drug, so they were only aware of material published after 1965.[44] Pre-1966 evidence about the risks associated with inhaled hexamethonium was available in *The Cochrane Library* and other sources.[45] Researchers should be more ready to call on the skills of information scientists to help them avoid embarking on ill-conceived or frankly unnecessary research.[46]

What about Bradford Hill's fourth question – 'What does it mean, anyway?' Getting a reliable answer to this question is of great importance to consumers of research evidence because it is the 'bottom line', and so may influence choices, practices or policies. Lord Rayleigh, professor of physics at Cambridge University, had something to say about this in his presidential address to the 54th Meeting of the British Association for the Advancement of Science in Montreal in 1884:

> *If, as is sometimes supposed, science consisted in nothing but the laborious accumulation of facts, it would soon come to a standstill, crushed, as it were, under its own weight ... The work which deserves, but I am afraid does not always receive, the most credit is that in which discovery and explanation go hand in hand, in which not only are new facts presented, but their relation to old ones is pointed out.*[47]

More than a century later, researchers and journal editors in most fields of scientific investigation have not taken his admonition seriously. Authors have too rarely assessed systematically the relation between 'new facts' and 'old facts' in the discussion sections of their reports of new research.[48] Even reports of randomised trials published in the most prestigious general medical journals leave Bradford Hill's fourth question inadequately addressed (Table 2.1).

All new research – whether basic or applied – should be designed in the light of scientifically defensible syntheses of existing research evidence, and reported setting

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

Classification	May 1997 (*n* = 26)	May 2001 (*n* = 33)	May 2005 (*n* = 18)
First trial addressing the question	1	3	3
Contained an updated systematic review integrating the new results	2	0	0
Discussed a previous review but did not attempt to integrate the new results	4	3	5
No apparent systematic attempt to set the results in the context of other trials	19	27	10

**The Lancet, New England Journal of Medicine, BMJ, JAMA* and *Annals of Internal Medicine*. Data from Clarke and co-workers.[49–51]

Table 2.1: **Classification of discussion sections in reports of randomised controlled trials published in May 1997, May 2001 and May 2005 in five general medical journals***

the new evidence 'in the light of the totality of the available evidence'.[26] This makes clear to readers what contribution – if any – new studies have made to knowledge.

As an illustration of the failure of researchers to take account of relevant evidence when designing and reporting new research, consider an analysis of reports of trials assessing the effect of aprotinin on the use of perioperative blood transfusion.[52] The first trial was reported in *The Lancet* in 1987, and showed a dramatically lower use of blood transfusions among patients who had received aprotinin than among control patients.[53] This difference was confirmed in 14 trials done over the subsequent 5 years. Yet, over the subsequent decade, a further 49 trials were reported. Figure 2.1 shows an analysis of the extent to which the authors of the reports of the 64 trials published by 2002 had cited relevant earlier trials. The shocking message in the figure is that, between 1987 and 2002, the proportion of relevant previous reports cited in successive reports of trials of aprotinin fell from a high of 33% to only 10% among the most recent reports. Furthermore, only seven of 44 subsequent reports referenced the report of the largest trial (which was 28 times larger than the median trial size); and most of the reports failed to reference either of the systematic reviews of relevant trials published in 1994 and 1997, respectively.

The human consequences of biomedicine's failure to cumulate evidence systematically

A pioneering study reported by Antman et al. in the *Journal of the American Medical Association* in 1992[28] used the technique of cumulative meta-analysis of randomised trials to show how much more would have been confidently known about the effects of treatments for myocardial infarction had successive trials set new results in the context of up-to-date systematic reviews of other relevant evidence. Figure 1 in their paper (reproduced here in Figure 2.2) shows this, using data on the effects of beta-blockade for secondary prevention of myocardial infarction.

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

The consequences of failing to fully use all relevant evidence 2

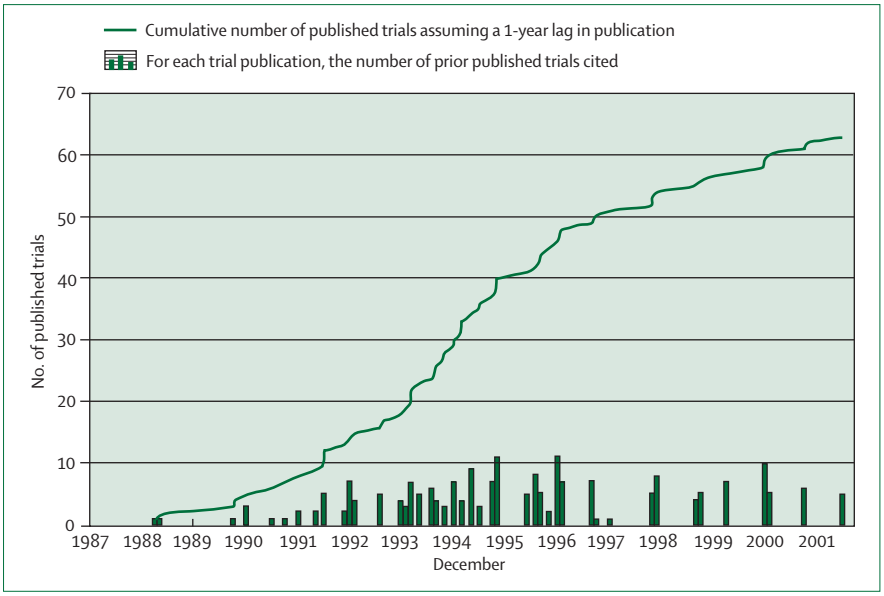


Figure 2.1: **Citations of relevant prior publications in 64 reports of randomised trials of aprotinin, 1988–2002. (Reproduced from Fergusson et al.[52] by permission of the Society for Clinical Trials)**

Figure 2.2(a) shows a now-familiar 'forest plot' presenting the results of 17 controlled trials comparing rates of death among patients receiving a beta-blocker drug with death rates among patients allocated to control. Along with the date of publication and number of participants in each trial, the results are represented by horizontal lines (confidence intervals) – the shorter the line, the more certain the result, reflecting the greater numbers of outcome events experienced by patients participating in that trial. The vertical line indicates the position around which the horizontal lines would cluster if the two treatments compared in the trials had similar effects. If a horizontal line touches or crosses the vertical line, it means that that particular trial found no statistically significant difference between the outcome of the drug group and that of the control group. Only two of these 17 trials individually yielded statistically significant estimates, which suggested that beta-blockers after myocardial infarction reduce mortality. However, it is not difficult to see that the horizontal lines tend to lie to the left of the vertical line. And indeed, when data derived from all 20 138 patients who participated in these studies is taken into account using meta-analysis, clear evidence emerges (shown by the bottom line) that these drugs have important beneficial effects.

Figure 2.2(b), which has a different horizontal scale to make it easier to see the confidence intervals, presents an analysis based on the same 20 138 patients in the same 17 trials, but arrayed in a different way – as a cumulative meta-analysis. This plot shows how estimates of the effects of beta-blockers would have looked had they been updated in the discussion sections of successive reports of each new trial. It shows that, had each of the successive reports published between 1972 and 1981 done this, chance would have been

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

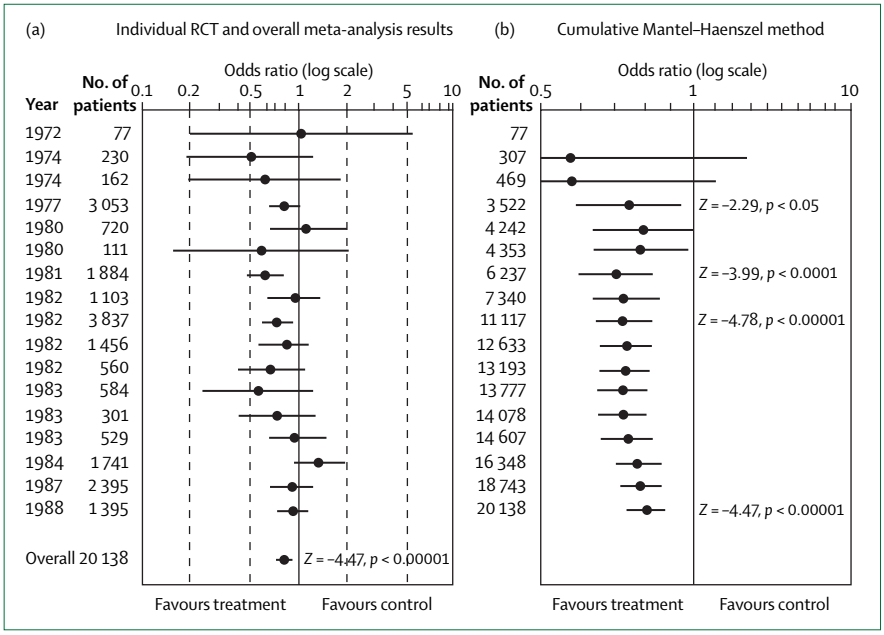


Figure 2.2: **(a) Standard and (b) cumulative meta-analyses of the results of 17 randomised control trials of the effects of oral beta-blockers for secondary prevention of mortality in patients surviving a myocardial infarction. (Reproduced from Antman et al.,[28] with permission. Copyright American Medical Association © 1992. All rights reserved)**

ruled out very confidently by 1981, after only six trials involving a total of 6237 patients had been reported. Although it may have been reasonable to do a few more placebo-controlled trials of these drugs, it is scientifically and ethically highly questionable whether nearly 14 000 patients needed to participate in further such studies.

Beta-blockade after myocardial infarction was just one of a number of treatments for myocardial infarction which Antman et al. evaluated in this way. For example, although strong evidence about the beneficial effects of thrombolytic therapy in myocardial infarction could, in principle, have been available by the mid-1970s, and a systematic review of this evidence was published in the *New England Journal of Medicine* in 1982,[54] the beneficial effects of thrombolysis were not mentioned at all in most textbooks until the late 1980s, and even when thrombolysis was mentioned, it was sometimes dismissed as unproven.[55]

The analyses done by Antman et al.[28] showed how science's failure to cumulate evidence scientifically had led to lethally incorrect advice in textbooks between 1960 and 1990: not only had advice on some life-saving therapies been delayed for more than a decade, but other treatments had been recommended long after controlled research had shown them to be harmful. An example of delayed recognition of lethal effects

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

concerns prophylactic use of antiarrhythmic drugs in myocardial infarction. Hundreds of thousands of premature deaths would have been prevented if the hazards had been recognised earlier. Although there was evidence from the 1970s showing that these drugs did indeed reduce arrhythmias,[56] doubts about their effectiveness in reducing death were first raised in a systematic review of 14 randomised trials published in the early 1980s.[57] By the late 1980s two further systematic reviews had confirmed not only that there was no evidence that these drugs had beneficial effects on mortality, but that they were probably lethal.[58, 59] By the time that their lethal potential had become generally accepted in the early 1990s, more than 50 randomised controlled trials involving 23 229 patients had been reported.[60] If each new report of the many randomised trials of a class 1 antiarrhythmic drug had set new results in the context of a systematic review of the results of all previous trials – in other words, if scientists had cumulated evidence scientifically – the lethal potential of these drugs could have been widely recognised a decade earlier. As already noted, it has been estimated that, at the peak of their use in the late 1980s, these drugs were causing tens of thousands of deaths every year in the USA[14] – comparable annual numbers of deaths to the *total* number of Americans who died in the Vietnam war.

There are now many examples of the consequences for patients of failure to cumulate evidence scientifically. It is obviously particularly important to identify treatments with harmful effects more efficiently, for example, postoperative radiotherapy in non-small-cell lung cancer,[61] and treatments with toxic effects which confer no advantage compared with less toxic alternatives, as systematic reviews have demonstrated in respect of treatments for ovarian cancer.[62]

A further illustration of the way in which beneficial effects of treatments will be missed unless evidence is cumulated scientifically concerns the long delay in recognising one of the most effective and cost-effective interventions for preventing neonatal morbidity and mortality.[63–68] In 1969, Liggins[69] reported his observation that ewes in whom he had induced labour prematurely with steroids had given birth to lambs who, unexpectedly, had air in their lungs. Liggins and Howie promptly began a randomised, placebo-controlled trial to assess the effect of giving a short course of corticosteroids to pregnant women who were expected to deliver prematurely. Within 3 years of the report on neonatal lambs they had reported a statistically significant lower infant morbidity and mortality in infants born to mothers who had received steroids compared with those who had received placebo.[70]

Many additional, smaller trials were done, and systematic reviews of these done in the 1980s[63, 64] made clear the important beneficial effects of prenatal corticosteroids (the results of the first seven trials form the basis of the logo of The Cochrane Collaboration; Figure 2.3). **{AQ_05}** More than 20 years after the initial trial had been published, in the light of continuing uncertainty among practitioners about the value of the treatment, the National Institutes of Health convened a consensus conference to assess the evidence. A cumulative meta-analysis prepared for the conference showed that, had scientists cumulated evidence systematically in reports of successive trials, there would have been little room for doubt about the importance of the treatment for at least the previous 15 years.[71] An account of the missed opportunities to reduce neonatal morbid-

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

Figure 2.3: **the Cochrane Collaboration logo (Reproduced with permission of The Cochrane Collaboration)**

ity and mortality (and the costs of neonatal intensive care) is available in the proceedings of a Wellcome Trust Witness Seminar.[72]

Some medical interventions with plausible effects have been promulgated very widely without evidence from randomised trials that they are likely to do more good than harm. In these circumstances, if systematic reviews have made explicit the absence of relevant evidence from randomised trials, it is clearly important to promote the needed trials, as was done eventually to assess the long-term effects of hormone replacement therapy, for example. In some circumstances, however, randomised trials of sufficient size may be particularly difficult to organise, and it may be necessary to rely on systematic reviews of evidence derived from the studies of the best available observational data.

One such example relates to advice to put infants to sleep on their fronts (prone) in the belief that this would reduce the risk of death from choking. Based on untested logic, this advice was promulgated between the mid-1950s and the late-1980s,[73] notably by one of the most influential paediatricians during this era, Dr Benjamin Spock, whose book *Baby and Child Care* had become a multimillion-copy best seller. From the 1958 edition of the book he recommended front sleeping because, he suggested, babies sleeping on their backs are more likely to choke if they vomit.

Partly because of the dramatic impact of the 'Back to Sleep' campaigns launched in the late 1980s and early 1990s, we now know that influential earlier promulgation of front sleeping advice by Spock and others led to tens of thousands of avoidable cot deaths.[73] It is an example of the dangers of introducing new practices on the basis of theory unsupported by evidence. In addition, however, it is a further example of the consequences of failure of researchers to develop a cumulative science. As Gilbert et al. have shown,[73] advice promulgated for nearly half a century to put infants to sleep on the front was contrary to evidence available from 1970 that this was likely to be harmful. They estimate that systematic review of the evidence as it accumulated (Figure 2.4) would have led to earlier recognition of the risks of sleeping on the front and might have prevented over 10 000 infant deaths in the UK and at least 50 000 in Europe, the USA and Australasia.

Although the above examples of failure to cumulate evidence systematically using methods to reduce biases and the play of chance are all from clinical and epidemiological research, it would be wrong to leave the impression that it is only within these spheres that there should be cause for concern about the failure of scientists to cumu-

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

The consequences of failing to fully use all relevant evidence 2

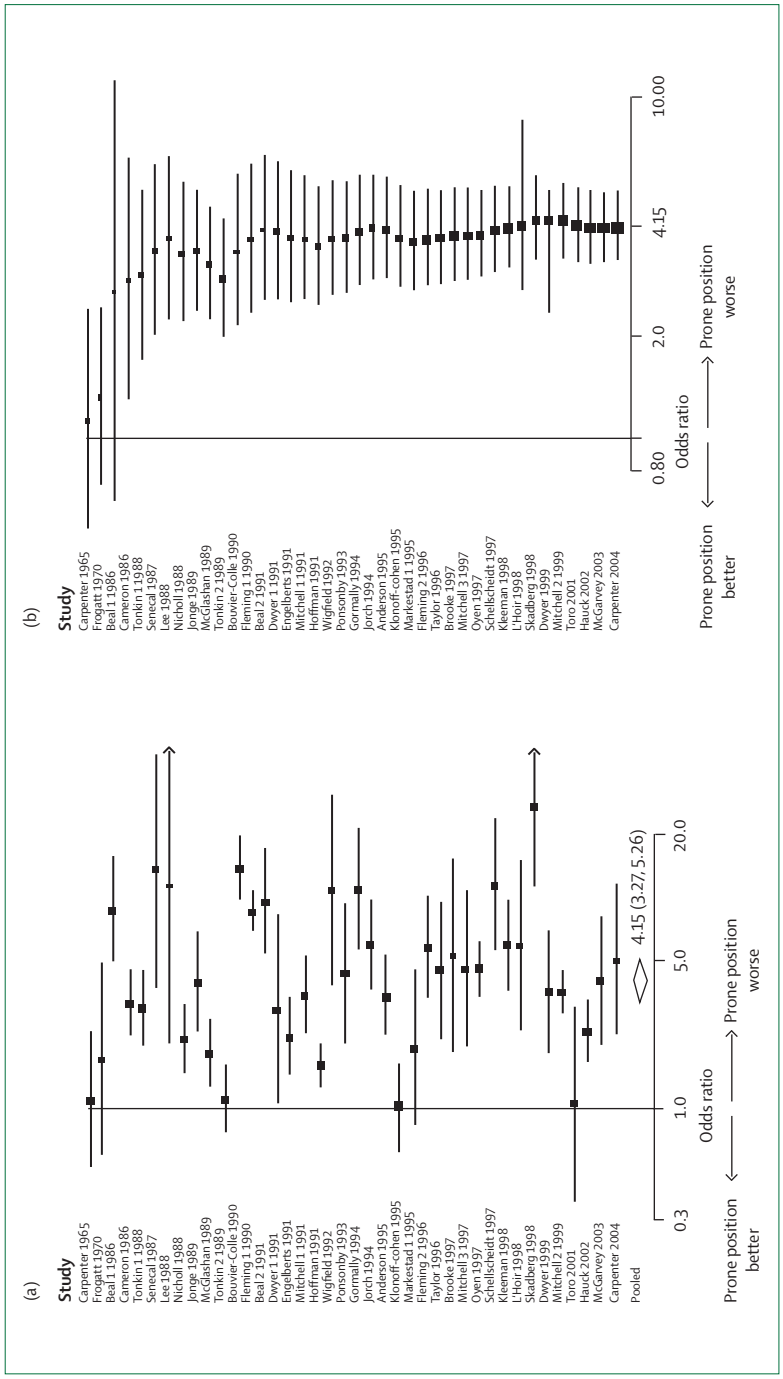


Figure 2.4: (a) Odds ratios and pooled odds ratio and (b) cumulative odds ratios for front versus non-front sleeping position in comparisons of cases of sudden infant death syndrome (SIDS) and controls (Reproduced from Gilbert et al.,[73] with permission of Oxford University Press and the International Epidemiological Association)

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

late scientifically. On the contrary, this issue is only now beginning to be addressed in preclinical research.

Failure to cumulate evidence systematically from experiments using animal models of human disease, for example, is not only bad science, it too can have adverse consequences for patients. Because it had been shown that calcium influx into areas of brain affected by acute ischaemic stroke led to cell death, it was thought that drugs that blocked calcium might protect the brain from damage. Based on selected reports of the early animal evidence, randomised clinical trials were begun. However, a systematic review of the evidence generated from randomised trials involving nearly 8000 patients found no evidence of any beneficial effects of nimodipine.[74] This finding prompted the authors to investigate the evidence from animal experiments. They concluded:

> *The results of this review did not show convincing evidence to substantiate the decision to perform trials with nimodipine in large numbers of patients. There were no differences between the results of the animal experiments and clinical studies. Surprisingly, we found that animal experiments and clinical studies ran simultaneously.*[75]

Had researchers using animal models of acute ischaemic stroke cumulated emerging evidence about the effects of nimodipine scientifically, it seems unlikely that it would have been judged worth inviting thousands of patients to participate in resource-intensive clinical trials of the drug. It is even more bizarre that animal experiments of nimodipine continued to be done for a considerable time after clinical trials had been initiated.[75]

Worrying examples of this kind have prompted challenges to animal researchers to be more systematic in developing the evidence base for their work.[76, 77] Systematic reviews of animal studies of six treatments that had been shown in systematic reviews of clinical trials to be either beneficial or to be harmful revealed a mixed picture of concordance and contradiction.[78] Formal assessment of the quality of animal research is beginning to yield some unsettling evidence.[79, 80] In 2002, for example, Bebarta et al.[81] showed that animal studies that had not used randomisation and blinding were more likely to have reported a difference between study groups than were studies that had used these methods to control biases – a comparable finding to that shown several years earlier in similar research analysing trials involving patients.[82] Assessment of the scientific quality of reviews of animal research also reveals serious deficiencies:[83] a large majority (i) did not specify a testable hypothesis, (ii) applied language restrictions, and (iii) failed to assess the possibility of publication bias or explore other reasons for heterogeneity. In only half of the reviews was the validity of component studies assessed.

A report on the ethics of research involving animals published by the Nuffield Council on Bioethics[84] in 2005 leaves little room for doubt that a serious problem exists, and that it must be addressed. The report notes the relative scarcity of systematic reviews and meta-analyses to evaluate more fully the predictability and transferability of animal models, and recommends that:

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

> *Since the scientific evaluation of animal research is fundamental to the cost–benefit assessment of any research, we recommend that the Home Office, in collaboration with major funders of such research such as the Wellcome Trust, the MRC, the BBSRC, animal protection groups and industry associations such as the ABPI, should consider ways of funding and carrying out these reviews.*[84]

At the time of writing, only the NHS Research and Development Programme, which does not fund research involving animals but has an unparalleled track record of supporting systematic reviews, has taken steps in response to this recommendation.[78]

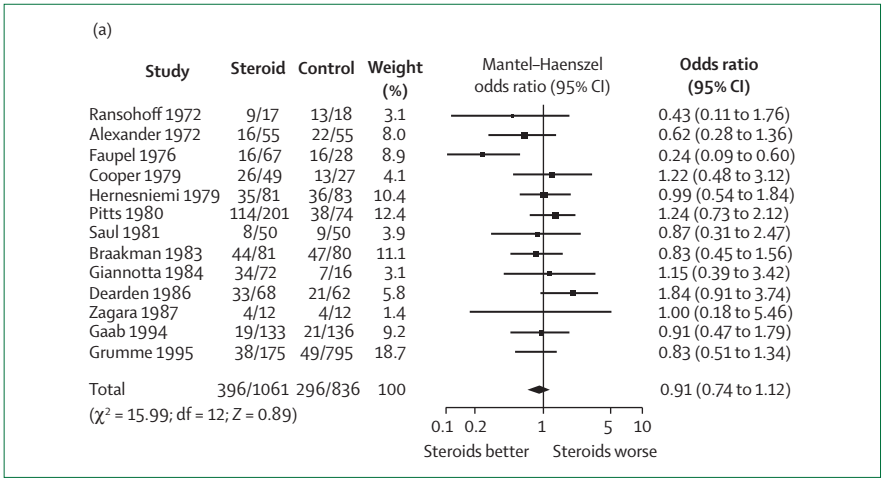
Meeting the needs of clinicians and patients more effectively

When will readers of reports of clinical research be able to expect to find answers to the fourth of the questions posed by Bradford Hill in 1965 – 'What does it mean anyway?'[25] It is not that there are no examples showing how new results can be set systematically in the context of other evidence. Some of these examples go back centuries. Indeed, in many respects James Lind did a better job of this in his book on scurvy published in 1753 than many researchers and journals do today.[85] Although they remain rare, however, there are examples of discussion sections of clinical trials containing systematic reviews in more recent times (see, e.g., Saunders et al.[86]). In days when space was at a premium in printed journals, editors complained that this would use too many of the journal's pages. For example, in 1986, an editorial in *The Lancet* referred to the meta-analysis presented in the discussion section of the 10-page report of the ISIS-1 trial as a 'lengthy tailpiece'. While acknowledging that 'there is a good case for such analyses', the editorial went on to make clear that 'if anyone suggests that they should become a regular feature of clinical trial reports *The Lancet* will lead the opposition'.[87]

As was pointed out in response at the time, however, these problems are not insuperable in an age of electronic publishing.[88, 89] The expectation that a report of a new randomised trial will begin by reference to the systematic review(s) that prompted the investigators to embark on the study, and end by setting the new results in the context of an up-to-date systematic review of trials relevant at the time of publication, does not imply that the introductory and discussion sections of every report of a randomised trial should contain a full account of the material, methods and findings of the reviews. The technology already exists to link to relevant, up-to-date systematic reviews published elsewhere.[90–93]

As an example of the kind of process and report of research needed, consider recent research on the effects of systemic steroids given to people with acute traumatic brain injury. A systematic review of existing evidence – published and unpublished – was done in the late 1990s as part of a programme of reviews assessing evidence of the effects of interventions in injured patients. As shown in Figure 2.5(a), the review revealed uncertainty about whether this treatment did more good than harm,[94] and this uncertainty was reflected in variations in the extent to which systematic steroids were

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.



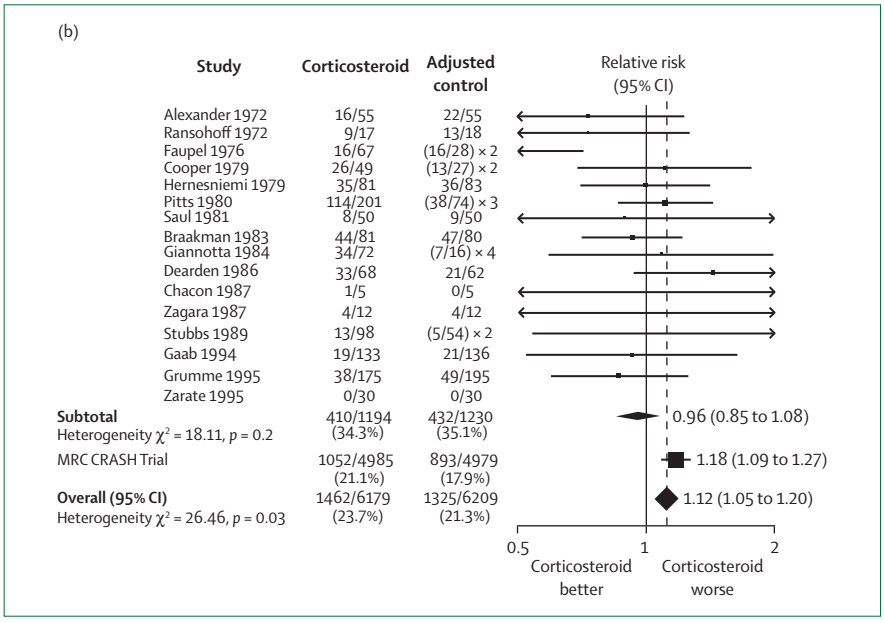


Figure 2.5: **Systematic reviews and meta-analyses of the effect of systemic corticosteroids in acute traumatic brain injury in 1997, 2004 and 2005. (a) Reproduced from Alderson and Roberts,[94] with permission of the *British Medical Journal*. (b) Reprinted from CRASH Trial Collaborators,[95] copyright 2001, with permission from Elsevier. (c) Reproduced from Alderson and Roberts,[96] with permission of The Cochrane Collaboration.**

used in clinical practice. Because this uncertainty related to a problem of global significance, a proposal for a large, multinational, randomised trial to address the uncertainty was submitted to the British Medical Research Council. The proposal met the Council's

The consequences of failing to fully use all relevant evidence 2

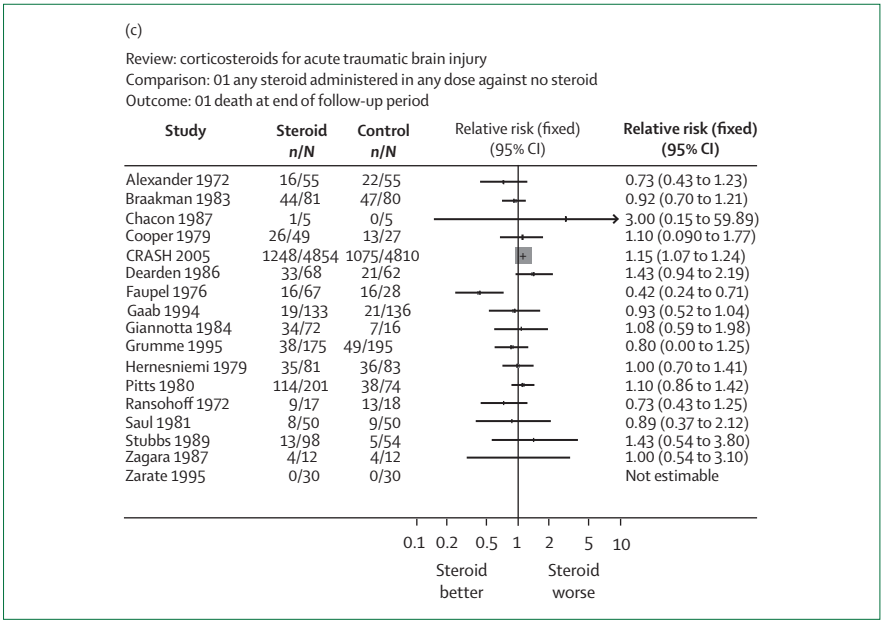


Figure 2.5: Continued

requirement[32] that the applicants should show, by reference to systematic reviews, why the trial was needed. Funding was agreed, and the results of the trial were subsequently reported in *The Lancet*.[95] This commissioning process and the report of the trial is exemplary because: (i) it refers to current uncertainty about the effects of a treatment, as manifested in the systematic review of all the existing evidence (published and unpublished), and in variations in clinical practice; (ii) the systematic review was a prerequisite required by the funding agency approached to support further research to address the uncertainty; (iii) the introduction of the trial report provides this background information; (iv) the discussion section of the report sets the new evidence in the context of an updated systematic review of all the evidence (Figure 2.5(b)), thus providing readers with all the evidence needed for action to prevent thousands of iatrogenic deaths; and (v) the electronically published version of the relevant review in the Cochrane Library was updated promptly to take account of the new evidence[96] (Figure 2.5(c)).

After giving examples of the adverse human consequences of the failure of biomedical scientists to cumulate scientifically in this way, editors at *The Lancet* wrote as follows:

> *In recognition that journal editors have a key part to play in ensuring that published research is presented in a way that clearly illustrates why it was necessary and what impact a particular trial has on the existing state of knowledge,* The Lancet *has decided to update its policies in this area. From August, 2005, we will require*

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

> *authors of clinical trials submitted to* The Lancet *to include a clear summary of previous research findings, and to explain how their trial's findings affect this summary. The relation between existing and new evidence should be illustrated by direct reference to an existing systematic review and meta-analysis. When a systematic review or meta-analysis does not exist, authors are encouraged to do their own. If this is not possible, authors should describe in a structured way the qualitative association between their research and previous findings.*[97]

The Lancet report of the ESPRIT trial is a welcome example of this policy being implemented in practice. The introduction in the report explained that the rationale for the trial had been uncertainty, made clear in systematic reviews of existing data, about the secondary preventive value of combined dipyridamole and aspirin in cerebral ischaemia of arterial origin. The Discussion section of the new trial incorporated the new results in an updated systematic review, enabling the authors to conclude with the clinically important statement that: 'The ESPRIT results, combined with the results of previous trials, provide sufficient evidence to prefer the combination regimen of aspirin plus dipyridamole over aspirin alone'.[98]

These are important examples of what is needed, but if the fullest possible use is to be made of research evidence to inform treatment decisions in individual patients, more analyses based on individual patient data are needed, such as those pioneered by the Antiplatelet Trialists' Collaboration,[18] the Early Breast Cancer Trialists' Collaborative Group[99] and the Advanced Ovarian Cancer Trialists' Group.[100] These collaborative reanalyses provide the additional flexibility needed to explore and confirm some of the real modifiers of treatment effects that can help clinicians and patients to individualise treatment decisions.

Glass, who coined the word meta-analysis in 1976,[101] observed 25 years later that 'meta-analysis was created out of the need to extract useful information from the cryptic records of inferential data analyses in the abbreviated reports of research in journals and other printed sources'.[102] He had not imagined that it would still be necessary, thirty years later, to rely on such sources (G. V. Glass, personal communication, **{AQ_02}**). If understanding how to use the results of randomised trials to inform decisions in personalised medicine really was the priority it should be, research synthesis should, by now, have become based on publicly accessible archives of raw data.[91] Indeed, responding to growing evidence that the current model of publishing clinical trials is hopelessly perverted, Smith and Roberts[103] have suggested it should be abandoned in favour a radical new approach based on cumulating data in publicly accessible databases. As they and others[104] have made clear, however, these changes will require less selfish and secretive attitudes within the biomedical research community.

Conclusion

As Lord Balfour's remarks in 1884 make clear, the concept of research synthesis is not new. In addition to medicine, there have been systematic reviews in such diverse topics

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

as advertising, agriculture, archaeology, astronomy, biology, chemistry, criminology, ecology, education, entomology, law, manufacturing, parapsychology, psychology, public policy, zoology and even eyewitness accounts of the Indian rope trick.[105] Reports of systematic reviews are cited more frequently than reports of other types of clinical research, including randomised trials, and this trend has been becoming stronger.[106] Furthermore, systematic reviews are likely to be a highly cost-effective form of research;[107] in particular, where there are several underpowered trials a systematic review can provide the needed power at a fraction of the cost of a 'mega-trial'.

Yet recognition that science is cumulative and that scientists need to cumulate scientifically using systematic reviews is still not accepted by many of the influential people who occupy the towering heights of biomedical academia.[108, 109] Some have felt the need to lampoon the very notion of systematic reviews – 'meta-analysis, shmeta-analysis',[110] 'statistical alchemy for the 21st century'.[111] Indeed, one senior academic has written: 'The idea of a systematic review is a nonsense, and the sooner those advocates of it are tried at the International Court of Human Rights at the Hague (or worse still, sent for counselling), the better'.[112]

These dismissive attitudes can have dangerous consequences. A systematic review and meta-analysis which had found no support for the widely promoted claim that postmenopausal hormone therapy prevents cardiovascular events (113) prompted the dean of a medical school to comment in a *BMJ* editorial: 'The correspondence columns this week will also reinforce readers wariness of meta-analysis ... For one I shall continue to tell my patients that hormone replacement therapy is likely to help prevent coronary disease'.[114] Attitudes such as this can have important adverse consequences for patients, as illustrated with the evidence presented in this chapter, and because they can lead to such outcomes, those who wield power within the biomedical research community should either publicly defend their failure to take effective action, or act more forcefully to change the unacceptable state of affairs described in this chapter and elsewhere.[109, 115]

Editors at *The Lancet* commented on the situation as follows:

> *Unnecessary and badly presented clinical research injures volunteers and patients as surely as any other form of bad medicine, as well as wasting resources and abusing the trust placed in investigators by their trial participants. Those who say that systematic reviews and meta-analyses are not 'proper research' are wrong; it is clinical trials done in the absence of such reviews and meta-analyses that are improper, scientifically and ethically. Investigators and organisations who undertake and coordinate reviews and meta-analyses now need the funding and recognition they deserve if public trust in biomedical research is to be maintained and resources used in an effective way.*[97]

The changes needed undoubtedly pose a challenge to researchers who are successful under the existing system. It requires them to undertake systematic reviews before applying for funds to do additional primary research, and to update those reviews when

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

they come to publish the results of new studies, and to collaborate with others doing related studies. This will be a new discipline for many of them, so it is easy to understand why they would oppose the changes needed.[116] It seems probable, however, that those who tacitly defend the status quo will increasingly have to defend their positions to the public. Articles in the lay press[117, 118] and popular books written for a lay readership[119] are beginning to draw attention to the issues discussed in this chapter. If research funders, academia, researchers, research ethics committees and scientific journals do not start to deal with the problems outlined in this chapter, the public seems likely to ask increasingly why they are acquiescing in what is clearly an indefensible state of affairs.

Systematic reviews are not a panacea: their results are a necessary but insufficient basis for informing policies in healthcare and treatment decisions in individual patients.[120] Furthermore, an increasing body of empirical research is beginning to explore the reasons why separate systematic reviews purportedly addressing the same question come to differing conclusions. Sometimes, for example, this may because the questions addressed are different; sometimes because the searches for potentially eligible studies differ in thoroughness; sometimes because of variations in inclusion criteria, or access to unpublished information; and sometimes because of differing spins put on essentially the same body of empirical evidence. The science of research synthesis is still young, and that is one of the reasons why current methodological discussions and research, and the insights that are emerging, make it a fulfilling field to work in and to develop.

That said, clinicians and patients need to be able to draw on syntheses of research evidence conducted to the highest standards possible, and the analyses in these systematic reviews should be able to draw on *all* the relevant research evidence. To increase the likelihood of making progress towards individualising treatment decisions, research syntheses must increasingly be based on collaborative analyses using individual patient data, as illustrated nicely in a recent reanalysis of controlled trials of treatment for otitis media,[121] improving on an influential earlier systematic review.

The scientific and ethical consequences of failure to take sufficiently seriously research synthesis, biased underreporting of clinical research, and access to the most useful data are that patients (and the public more generally) suffer directly and indirectly; policymakers, practitioners and patients have inadequate information to guide their choices among alternatives for individual patients; and limited resources for healthcare and new research are used inefficiently.

Acknowledgements

I am grateful to Doug Altman, Amanda Burls, Luis-Gabriel Cuervo, Frank Davidoff, Matthias Egger, Paul Glasziou, Andrew Herxheimer, Chris Hyde, Tom Jefferson, David Mant, Tom Marshall, Drummond Rennie, Peter Rothwell, David Sackett, Josie Sandercock, David Schriger, Loes van Bokhoven, Floris van de Laar and Simon Wessely for helpful discussion of the ideas and material presented in this chapter.

References

1 Poisson, Dulong, Larrey et Double. Recherches de statistique sur l'affection calculeuse, par M. le docteur Civiale. *Comptes rendus hebdomadaires des séances de l'Académie de Sciences.* Paris: Bachelier, 1835, pp 167–177. [Statistical research on conditions caused by calculi by Doctor Civiale. Translated and reprinted in *Int J Epidemiol* 2001; **30:** 1246–49.]

2 Mant D. Can randomised trials inform clinical decisions about individual patients? *Lancet* 1999; **353:** 743–46.

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

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

4 Chalmers I. What do I want from health research and researchers when I am a patient? *BMJ* 1995; **310:** 1315–1318.

5 Baker SG, Kaprio J. Common susceptibility genes for cancer: search for the end of the rainbow. *BMJ* 2006; **332:** 1150–52.

6 New B. Defining a package of healthcare services the NHS is responsible for. The case for. *BMJ* 1997; **314:** 503–05.

7 Klein R. Defining a package of healthcare services the NHS is responsible for. The case against. *BMJ* 1997; **314:** 505–09.

8 Oxman AD, Guyatt GF. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992; **116:** 78–84.

9 Glasziou P, Irwig L. An evidence-based approach to individualising patient treatment. *BMJ* 1995; **311:** 1356–59.

10 Bracken M. Genomic epidemiology of complex disease: the need for an evidence-based electronic approach to research synthesis. *Am J Epidemiol* 2005; **162:** 1–5.

11 Chalmers I. A patient's attitude to the use of research evidence to guide individual choices and decisions in health care. *Clin Risk* 2000; **6:** 227–30.

12 Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 2005; **365:** 1348–53.

13 Cowley AJ, Skene A, Stainer K, Hampton JR. The effect of lorcainide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias. *Int J Cardiol* 1993; **40:** 161–66.

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

15 Chalmers I. From optimism to disillusion about commitment to transparency in the medico-industrial complex. *J R Soc Med* 2006; **99:** 337–41.

16 Jack A. Call to release human drug trial. *Financial Times*, 8 August 2006.

17 Global R&D performance metrics programme: industry success rates report. *CMR International*, May 2005, p 7.

18 Antiplatelet Trialists' Collaboration. Secondary prevention of vascular disease by prolonged anti-platelet treatment. *BMJ* 1988; **296:** 320–31.

19 Sacks FM, Pfeffer MA, Moye LA, et al. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators. *N Engl J Med* 1996; **335:** 1001–09.

20 Bonovas S, Filioussi K, Tsavaris N, Sitaras NM. Use of statins and breast cancer: a meta-analysis of seven randomized clinical trials and nine observational studies. *J Clin Oncol* 2005; **23:** 8606–12.

21 Last JM. *A Dictionary of Epidemiology*. Oxford: Oxford University Press, 2001.

22 Pearson K. Report of certain enteric fever inoculation statistics. *BMJ* 1904; **3:** 1245–46.

23 Glasziou P, Chalmers I, Rawlins M, McCulloch P. Detecting the signal of a treatment effect from the noise of prognosis: judging when randomized trials are essential. Submitted. **{AQ_03}**

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

25 Hill AB. Cited in: The reasons for writing. *BMJ* 1965; **2:** 870–72.

26 Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; **276:** 637–39.

27 Baum ML, Anish DS, Chalmers TC, et al. A survey of clinical trials of antibiotic prophylaxis in colon surgery: evidence against further use of no-treatment controls. *N Engl J Med* 1981; **305:** 795–99.

28 Antman EM, Lau J, Kupelnick B, et al. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *JAMA* 1992; **268:** 240–48.

29 Aspinall RL, Goodman NW. Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for postoperative nausea and vomiting: a review of published trials. *BMJ* 1995; **311:** 844–46.

30 Cooper NJ, Jones DR, Sutton AJ. The use of systematic reviews when designing studies. Clin Trials 2005; 2: 260–64.

31 Chalmers I, Matthews R. What are the implications of optimism bias in clinical research? *Lancet* 2006; **367:** 449–50.

32 O'Toole L. Using systematically synthesized evidence to inform the funding of new clinical trials – the UK Medical Research Council approach. Paper presented at the 6th International Cochrane Colloquium, Baltimore, MD, 22–26 October 1998.

33 Kramer BS, Wilentz J, Alexander D, et al. Getting it right: being smarter about clinical trials. *PLoS Med* 2006; **3:** e144.

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

34 Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 1996; **313:** 1390–93.

35 Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med* 1994; **331:** 384–89.

36 Smith AJ, Goodman NW. The hypertensive response to intubation. Do researchers acknowledge previous work? *Can J Anaesth* 1997; **44:** 9–13.

37. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA* 1990; **263:** 1392–95.

38 Chalmers I. Role of systematic reviews in detecting plagiarism: case of Asim Kurjak. *BMJ* 2006; **333:** 594–96.

39 Jefferson T, Deeks J. The use of systematic reviews for editorial peer reviewing: a population approach. In: Godlee F, Jefferson T, eds. *Peer Review in Health Sciences*. London: BMJ Books, 1999: 224–34.

40 Jefferson T, Shashok K. Journals: how to decide what's worth publishing. Nature 2003; 421: 209–10.

41 Chalmers I. Lessons for research ethics committees. *Lancet* 2002; **359:** 174.

42 Mann H, Djulbegovic B. Why comparisons must address genuine uncertainties. James Lind Library, 2005. Available at: http://www.jameslindlibrary.org/trial_records/21st_Century/djulbegovic/djulbegovic_commentary.html (accessed February 2007)

43 Savulescu J, Spriggs M. The hexamethonium asthma study and the death of a normal volunteer in research. *J Med Ethics* 2002; **28:** 3–4.

44 McLellan F. 1966 and all that – when is a literature search done? *Lancet* 2001; **358:** 646.

45 Clark O, Clark L, Djulbegovic B. Is clinical research still too haphazard? *Lancet* 2001; **358:** 1648.

46 Florance V, Davidoff F. The informationist: a new health profession? *Ann Intern Med* 2000; **132:** 996–98.

47 Rayleigh, The Right Hon Lord. Presidential address at the 54th Meeting of the British Association for the Advancement of Science, Montreal, August/September 1884. London: John Murray, 1885: 3–23.

48 Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Evaluation Health Professions* 2002; **25:** 12–37.

49 Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? *JAMA* 1998; **280:** 280–82.

50 Clarke M, Alderson P, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals. *JAMA* 2002; **287:** 2799–801.

51 Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. Submitted. **{AQ_04}**

52 Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: using clinical equipoise to stop the bleeding. *Clin Trials* 2005; **2:** 218–32.

53 Royston D, Bidstrup BP, Taylor KM, Sapsford RN. Effect of aprotinin on the need for blood transfusion after repeat open-heart surgery. *Lancet* 1987; **2:** 1289–91.

54 Stampfer MJ, Goldhaber SZ, Yusuf S, et al. Effect of intravenous streptokinase on acute myocardial infarction: pooled results from randomized trials. *N Engl J Med* 1982; **307:** 1180–82.

55 Pentecost A. Myocardial infarction. In: Weatherall DJ, Ledingham JGG, Warrell DA, eds. *Oxford Textbook of Medicine*, 2nd edn. Oxford: Oxford University Press, 1987: 13.173.

56 Sheridan DJ, Crawford L, Rawlins MD, Julian DG. Antiarrhythmic action of lignocaine in early myocardial infarction. Plasma levels after combined intramuscular and intravenous administration. *Lancet* 1977; **1:** 824–25.

57 Furberg CD. Effect of anti-arrhythmic drugs on mortality after myocardial infarction. Am J Cardiol 1983; 52: 32C–36C.

58 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:** 2694–98.

59 MacMahon S, Collins R, Peto R, et al. Effects of prophylactic lidocaine in suspected acute myocardial infarction. An overview of results from the randomized, controlled trials. *JAMA* 1988; **260:** 1910–16.

60 Teo KK, Yusuf S, Furberg CD. Effects of prophylactic anti-arrhythmic drug therapy in acute myocardial infarction. *JAMA* 1993; **270:** 1589–95.

61 PORT Meta-analysis Trialists Group. Post-operative radiotherapy in non-small cell lung cancer: systematic review and meta-analysis of individual patient data from nine-randomised controlled trials. *Lancet* 1998; **352:** 257–63.

62 Sandercock J, Parmar MKB, Torri V, Qian W. Chemotherapy for advanced ovarian cancer: paclitaxel, cisplatin and the evidence. *Br J Cancer* 2002; **87:** 815–24.

63 Crowley PA. Corticosteroids in pregnancy: the benefits outweigh the costs. *J Obstet Gynaecol* 1981; **1:** 147–50.

64 Crowley PA. Promoting pulmonary maturity. In: Chalmers I, Enkin M, Keirse MJNC, eds. *Effective Care in Pregnancy and Childbirth*. Oxford: Oxford University Press, 1989: 746–64.

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

The consequences of failing to fully use all relevant evidence 2

65 Crowley P, Chalmers I, Keirse MJNC. The effects of corticosteroid administration before preterm delivery: an overview of the evidence from controlled trials. *Br J Obstet Gynaecol* 1990; **97:** 11–25.

66 Mugford M, Piercy J, Chalmers I. Cost implications of different approaches to the prevention of respiratory distress syndrome. *Arch Dis Child* 1991; **66:** 757–64.

67 Crowley PA. Antenatal corticosteroid therapy: a meta-analysis of the randomized trials, 1972–1994. *Am J Obstet Gynaecol* 1995; **173:** 322–35.

68 Hanney S, Mugford M, Grant J, Buxton M. Assessing the benefits of health research: lessons from research into the use of antenatal corticosteroids for the prevention of neonatal respiratory distress syndrome. *Soc Sci Med* 2005; **60:** 937–47.

69 Liggins GC. Premature delivery of lambs infused with glucocorticoids. *J Endocrinol* 1969; **45:** 515–23.

70 Liggins GC, Howie RN. A controlled trial of antepartum glucorticoid treatment for the prevention of the respiratory distress syndrome in premature infants. *Pediatrics* 1972; **50:** 515–25.

71 Sinclair JC. Meta-analysis of randomized trials antenatal corticosteroid therapy for the prevention of respiratory distress syndrome: discussion. *Am J Obstet Gynaecol* 1995; **173:** 335–44.

72 Reynolds L, Tansey EM. Prenatal *Corticosteroids for Reducing Morbidity and Mortality after Preterm Birth*. London: Wellcome Trust, 2005.

73 Gilbert R, Salanti G, Harden M, See S. Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of clinicians' recommendations from 1940–2000. *Int J Epidemiol* 2005; **34:** 74–87.

74 Horn J, Limburg M. Calcium antagonists for acute ischemic stroke (Cochrane Review). In: *The Cochrane Library*, Issue 3. Oxford: Update Software, 2001.

75 Horn J, de Haan RJ, Vermeulen M, et al. Nimodipine in animal model experiments of focal cerebral ischaemia: a systematic review. *Stroke* 2001; **32:** 2433–38.

76 Sandercock P, Roberts I. Systematic reviews of animal experiments. *Lancet* 2002; **2:** 586.

77 Pound P, Ebrahim S, Sandercock P, et al. Where is the evidence that animal research benefits humans? *BMJ* 2004; **328:** 514–17.

78 Perel P, Roberts I, Sena E, et al. Testing Treatments on Animals: Relevance to Humans. Final Report, May 2006. Available at: http://www.pcpoh.bham.ac.uk/publichealth/nccrm/publications.htm (accessed February 2007).

79 Khan KS, Mignini L. Surveying the literature from animal experiments: avoidance of bias is objective of systematic review, not meta-analysis. *BMJ* 2005; **331:** 110–11.

80 Macleod MR, Ebrahim S, Roberts I. Surveying the literature from animal experiments: systematic reviews and meta-analysis are important contributions. *BMJ* 2005; **331:** 110.

81 Bebarta V, Luyten D, Heard K. Emergency medicine animal research: does use of randomization and blinding affect the results? *Academic Emerg Med* 2003; **10:** 684–87.

82 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.

83 Mignini LE, Khan KS. Methodological quality of systematic reviews of animal studies: a survey of reviews of basic research. *BMC Med Res Methodol* 2006; **6:** 10 [DOI: 10.11186/1471–2288/6/10].

84 Nuffield Council on Bioethics. *The Ethics of Research Involving Animals*. London: Nuffield Foundation, 2005.

85 Milne I, Chalmers I. Documenting the evidence: the case of scurvy. Bull WHO 2004; 82: 791–92.

86 Saunders MC, Dick JS, Brown IMcL, et al. The effects of hospital admission for bed rest on the duration of twin pregnancy: a randomised trial. *Lancet* 1985; **2:** 793–95.

87 Editorial. Intravenous beta-blockade during acute myocardial infarction. *Lancet* 1986; **2:** 79–80.

88 Chalmers I. Electronic publications for updating controlled trial reviews. *Lancet* 1986; **2:** 287.

89 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: BMJ Books, 1991: 127–46.

90 Huth EJ. Quality in the electronic age. *Eur Sci Editing* 1997; **23:** 41–42.

91 Chalmers I, Altman DG. How can medical journals help prevent poor medical research? Some opportunities presented by electronic publishing. *Lancet* 1999; **353:** 490–93.

92 Chalmers I. Using systematic reviews and registers of ongoing trials for scientific and ethical trial design, monitoring, and reporting. In: Egger M, Davey Smith G, Altman D, eds. *Systematic Reviews in Health Care: Meta-Analysis in Context*, 2nd edn of Systematic Reviews. London: BMJ Books, 2001: 429–43.

93 Clarke M. Doing new research? Don't forget the old. Nobody should do a trial without reviewing what is known. *PLoS Med* 2004; **1:** 100–02.

94 Alderson P, Roberts I. Corticosteroids in acute traumatic brain injury: systematic review of randomised controlled trials. *BMJ* 1997; **314:** 1855–59.

Chalmers I. The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P, ed. Treating individuals: from randomised trials to personalised medicine. London: Lancet, 2007: 37-58.

95 CRASH Trial Collaborators. Effect of intravenous corticosteroids on death within 14 days in 10 008 adults with clinically significant head injury (MRC CRASH Trial): a randomised placebo-controlled trial. *Lancet* 2004; **364:** 1321–28.

96 Alderson P, Roberts I. Corticosteroids in acute traumatic brain injury. *Cochrane Database of Systematic Reviews*, 2005, Issue 1. Art. No.: CD000196 [DOI: 10.1002/14651858.CD000196.pub2].

97 Young C, Horton R. Putting clinical trials into context. *Lancet* 2005; **366:** 107–08.

98 ESPRIT Study Group. Aspirin plus dipyridamole versus aspirin alone after cerebral ischaemia of arterial origin (ESPRIT): randomised controlled trial. *Lancet* 2006; **367:** 1665–73.

99 Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer: and overview of 62 randomized trials among 28,896 women. *N Engl J Med* 1988; **319:** 1681–92.

100 Advanced Ovarian Cancer Trialists' Group. Chemotherapy in advanced ovarian cancer: an overview of randomised clinical trials. *BMJ* 1991; **303:** 884–93.

101 Glass GV. Primary, secondary and meta-analysis of research. *Educational Researcher* 1976; **10:** 3–8.

102 Glass GV (2001). Meta-analysis at 25. January 2000. Available at: http://glass.ed.asu.edu/gene/papers/meta25.html (accessed February 2007).

103 Smith R, Roberts I. Patient safety requires a new way to publish clinical trials. *PLoS Clin Trials* 2006, May, e6.

104 Vickers AJ. Whose data set is it anyway? Sharing raw data from randomized trials. *Trials* 2006; 7: 15 [DOI: 10.1186/1745-6215-7-15].

105 Petticrew M. Systematic reviews from astronomy to zoology: myths and misconceptions. *BMJ* 2001; **322:** 98–101.

106 Patsopoulos NA, Analatos AA, Ioannidis JP. Relative citation impact of various study designs in the health sciences. *JAMA* 2005; **293:** 2362–66.

107 Glasziou P, Djulbegovic B, Burls A. Are systematic reviews more cost-effective than randomized trials? *Lancet* 2006; **367:** 2057–58.

108 Alderson P, Gliddon L, Chalmers I. Academic recognition of critical appraisal and systematic reviews in British postgraduate medical education. *Med Educ* 2003; **37:** 386–87.

109 Chalmers I. Academia's failure to support systematic reviews. *Lancet* 2005; **365:** 469.

110 Shapiro S. Meta-analysis/shmeta-analysis. *Am J Epidemiol* 1994; **140:** 771–78.

111 Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 1995; **48:** 71–79.

112 Rees J. Two cultures? *J Am Acad Dermatol* 2002; **46:** 313–14.

113 Hemminki E, McPherson K. Impact of postmenopausal hormone therapy on cardiovascular events and cancer: pooled data from clinical trials. *BMJ* 1997; **315:** 149–53.

114 Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *BMJ* 1997; **315:** 617–19.

115 Chalmers I. The scandalous failure of science to cumulate evidence scientifically. *Clin Trials* 2005; **2:** 229–31.

116 Marshall T. Scientific knowledge in medicine: a new clinical epistemology? *J Evaluation Clin Pract* 1997; **3:** 133–38.

117 Brown D. Superfluous medical studies called into question. *The Washington Post*, 2 January 2006: A06.

118 Matthews R. No way to treat a patient: tens of thousands of people have been subjected to unnecessary drug trials. *New Scientist*, 9 July 2006: 19.

119 Evans I, Thornton H, Chalmers I. *Testing Treatments: Better Research for Better Healthcare*. London: British Library, 2006.

120 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.

121 Rovers M, Glasziou P, Appelman C, et al. Antibiotics for acute otitis media: a meta-analysis with individual patient data. *Lancet* 2006; **368:** 1429–35.