The end of medical history?

'The end of history' is the title of a book by Francis Fukuyama published in 1992. His thesis was that liberal democracy had emerged as a standard form of government throughout the world, displacing hereditary monarchy, fascism and communism. He suggested that liberal democracy represents the end of political evolution and the ultimate and permanent form of government, and its worldwide appearance thus represents 'the end of history'. He did not claim that major events do not still occur, but he argued that the broad sweep of historical progress seems to have come to an end; with a few notable and all too obvious exceptions the world has become politically bland and homogeneous.

Can we see the same process occurring in medicine? In a world where medical – or at least, biological – progress is so rapid, with publication numbers rising apparently exponentially, it may seem odd even to ask the question. But for the clinician the essentials of medicine are diagnosis and – above all – treatment. For the clinician developments in basic science may be crucial but they are not important; basic science is only important in so far as it leads to improvements in the management of patients. A clinician can make use of developments in basic science in the same way as he drives a car: the latest in engine technology may make for a smoother ride, but how that is achieved matters little.

If we take the narrow view, and assume that medicine as perceived by the clinician is essentially a matter of patient management, can we argue that history has come to end? A continuous stream of new treatments becomes available, with demonstrated benefits to patients, so surely history is still being made? I intend to make the case that with hindsight we can now see a point where the development of medical practice came to an end in the same way that socio-political development ended with the widespread acceptance of liberal democracy. Since that point in medical progress there have, of course, been important new developments but they have occurred on a background of homogeneous thought, and thus are not the stuff of which history is made. Perhaps we need to start by asking what, from the clinician's point of view, can be accepted as 'real' history.

What makes history?

If a clinician surveys the drugs he uses in his daily practice he can identify a few which he can say 'made history'. Probably none of these came into use in the last quarter century, and most will be much older.

We can make the point by thinking of two drugs, penicillin and digoxin. The development of penicillin makes a story enthralling enough to justify a place in any modern history book. A chance observation of the effect of a mould on a culture plate, ignored for a few years, was developed by a small team of scientists led by Howard (later Lord) Florey (Fig 1) at Oxford to the point where infections that were previously lethal could successfully be treated. The results of treating patients with this mould extract were so dramatic for the time that individual patients' temperature charts could be published in the Lancet. The excitement was captured twenty years later by the surgeon at the Radcliffe Infirmary, Oxford, who had been responsible for draining the abscesses of the first patient treated with penicillin (Fig 2). In the Oxford Medical School Gazette, Arthur Elliott Smith described how a policeman with septicemia responded to penicillin injections, and how his urine was collected and returned to the Dunn School of Pathology so that the excreted penicillin could be extracted and reused. After an initial and quite dramatic improvement the patient relapsed when the supply of penicillin could no longer be maintained, and he eventually died. Elliott Smith wrote:

'It should be remembered that these events in the Radcliffe Infirmary took place against a background of the most difficult stage of the war. The enemy bombing offensive was at its maximum during the autumn of 1940 and spring 1941. . . . On the night of April 8 medical staff ...'

There can be no debate that this was history. It fits exactly Winston Churchill's obituary of Neville Chamberlain:

'History with its flickering lamps stumbles along the trail of the past trying . . . to kindle with pale gleams the passion of former days.'

As a cardiologist it is perhaps inevitable that I should think of advances in cardiological treatment to support my theme, but any clinician would accept Withering's use of the foxglove for the treatment ofdropsy as being of major clinical significance, and justifying a key place in medical history. Withering realised that the foxglove must be the active component of a successful remedy used 'by an old Shropshire woman' and with a series of clinical experiments
he found out how to use it. He demonstrated that clinical
efficacy rapidly changed to harm unless the dose was care-
fully controlled, and he recognised that the foxglove was
only effective when the dropsy resulted from heart disease
and not from, for example, disease of the liver. In 1785 he
wrote:

If the pulse be feeble or intermitting, the countenance pale,
the lips livid, the skin cold, the swollen belly soft and fluctu-
ating, or the anasarcous limbs readily pitting under the
pressure of the fingers, we may expect the diuretic effects to
follow in a kindly manner.

The efficacy of penicillin was so obvious and so dramatic
that there has never been any need to question its impor-
tance. In the case of my second historical example, digitalis,
the story that has unfolded over the years has been less
clear-cut. For nearly 200 years digitalis, initially as an
extract of foxglove leaves, was the only effective remedy for
heart failure, but with the introduction of diuretics – first
the mercurials, then the thiazides, and then the loop diuret-
ics – digitalis became less important. Particularly in the
United Kingdom, cardiologists came to believe that digoxin
was only effective in so far as it controlled the ventricular
rate in patients with atrial fibrillation – those in whom ‘the
pulse was faint and intermitting’. Only in the last few years
has the precise role of digoxin been established for patients
with heart failure but whose hearts have remained in sinus
rhythm. We can now be reasonably certain that in such
patients treatment with digoxin neither prolongs nor
reduces survival; it reduces deaths from progressive heart
failure but increases deaths from myocardial infarction and
(presumed) arrhythmias, and the net effect is balanced.
Digoxin does, however, reduce hospitalisation for heart
failure.

How do we know this, and how have we ‘progressed’ from
the certainties of Withering and the universal acclaim for an
effective treatment for dropsy, to our present qualified
doubt about the value of digoxin therapy? The answer to
this question is, of course, that we now base our treatment
whenever possible on the results of clinical trials. Withering
based his writing on his memory of individual cases: while
he did not actually admit that his memory might be
selective for those patients who showed a dramatic benefit
from treatment, he did accept that his observations did not
represent the whole of his experience. He wrote:

The patients were such as applied at my house for advice
gratis. I cannot pretend to charge my memory with particular
cases, or particular effects, and I had not the leisure to make
notes. Upon the whole, however, it may be concluded
that the medicine was found useful, or I should not have
continued to employ it.

Contrast this with the publication on which current use
of digoxin is based. We do not even know its author: the
paper emanates from an anonymous ‘Digitalis Investigation
Group’; 6,800 patients at 302 hospitals in North America
were treated at random with either digoxin or dummy
tablets in addition to a wide range of conventional therapy

Fig 1. Dr Florey, later Lord Florey, ca 1940.

Fig 2. The first patient to be treated with penicillin. (Reproduced
from Ref 3 by permission of the Oxford Medical School
Gazette.)
now available. After a little more than three years there had been 1,181 deaths (34.8%) among the patients treated with digoxin, and 1,194 deaths (35.1%) among those given placebo tablets. The publication concludes with 95% confidence that the risk ratio, when digoxin was compared with placebo, was between 0.91 and 1.07. In other words, on the basis of all the effort that went into the study, the ‘real’ answer to the effect of digoxin could be anything between a reduction of the absolute risk of death by about 3% and a similar sized increase of risk. While the publication of the trial usually referred to as ‘DIG’ was undoubtedly both interesting and clinically important, it is difficult to see it as a historical monument equal in importance to the writings of William Withering.

To return to the analogy of Fukuyama’s ‘end of history’, the DIG trial is one of those undoubtedly significant events that have occurred in the now homogeneous world of medical advance on the basis of clinical trials. To find out if there was a point when our therapeutic climate developed this homogeneity, we need to review the history of the development of clinical trials. If we can identify a point beyond which no important new concepts of clinical trials appeared, we can reasonably talk about ‘the end of medical history’.

The development of clinical trials

The development of medical treatment was a fairly hit-and-miss affair until the eighteenth century. One of the major problems, of course, was an inability to diagnose diseases and to separate one from another. Presumably the persistent belief in the value of blood letting was based on observations of its dramatic benefit in left ventricular failure: a doctor who cannot differentiate between heart failure and pneumonia will not unnaturally assume that one treatment will be equally beneficial to both. Another major problem was identified in 1834 by Louis:

As to different methods of treatment, it is possible for us to assure ourselves of the superiority of one or other . . . by enquiring if the greater number of individuals have been cured by one means than another. Here it is necessary to count. And it is, in great part at least, because hitherto this method has not at all, or rarely been employed, that the science of therapeutics is so uncertain.

One disease that was well characterised and diagnosed with considerable accuracy in the eighteenth century was scurvy, the scourge of seamen. The treatment of scurvy formed the basis of what seems to have been the first real clinical trial.

James Lind was born in Edinburgh in 1716, and after serving his apprenticeship he became a member of the Incorporation of Surgeons and was a naval surgeon for nine years. During this time he saw two severe outbreaks of scurvy in the Channel Fleet, in one of which 80 men out of a crew of 350 were affected. In A treatise of the scurvy' (Fig 3), published in 1753, he described his experiment:

On the 20th of May 1747, I took twelve patients in the scurvy, on board the Salisbury at sea. The cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of the knees. They lay together in one place . . . and had one diet common to all. Two of these were ordered each a quart of cider a day, two others took 25 gutts of elixir vitriol 3 times a day upon an empty stomach. Two others took two spoonfuls of vinegar 3 times a day . . . Two of the worst patients were put under a course of seawater. Two others had each two oranges and one lemon given them every day . . . they continued but six days under this course, having consumed the quantity that could be spared. The two remaining patients took . . . an electuary recommended by a hospital surgeon. The consequence was that the most sudden and visible good effects were perceived from the use of oranges and lemons; one of those who had taken them being at the end of six days fit for duty, the other . . . was appointed to nurse the rest of the sick.

Lind dedicated his work to Admiral Lord Anson:

Who, as a just reward for the great and signal services done to the British Nation, does now preside over her Naval Affairs, the following Treatise is inscribed, with the greatest respect by His Lordship's most devoted and most obedient humble servant.

This seems to have been a curious thing to do considering that Anson had lost a record number of men to scurvy during his circumnavigation a few years earlier. Perhaps this is the first recorded example of a young doctor trying to gain advancement through publication; if so, Lind's
objective was achieved, for he was appointed as the first superintendent of the big new naval hospital, Haslar, in Portsmouth. It is, however, salutary to note that the routine introduction of lemon juice to the rations of sailors did not occur until two years after his death in 1794.

In the nineteenth and early twentieth centuries we can pick out a few therapeutic experiments which can be described as trials, even though it is difficult to find examples as elegant or with such a clear and dramatic result as Lind's study of the treatment of scurvy.

In 1870 Lister described the 'Effects of the antiseptic system of treatment upon the salubrity of a surgical hospital'69. This was not a comparative trial of two methods of treatment so much as an audit of results before and after the introduction of a new idea. Lister wrote:

'It is of the wards lately under my care were in the highest degree beneficial, converting them from some of the most unhealthy in the Kingdom into models of healthiness.

The 'antiseptic system' is not described in a way that would allow others to copy it easily, but the results were impressive. In the years 1864–66, 'before the antiseptic period', a total of 35 amputations were performed, with 16 deaths - or, as Lister put it, 1 death in every 2½ cases. In 1867–69, 'during the antiseptic period', there were 6 deaths after 40 amputations, or '1 death in every 6½ cases'. In this paper, published in the Lancet on 1 January 1870, Lister seemed to foresee the demands that the Editor of the Lancet would be making of his contributors a century later. 'These numbers are, no doubt', he wrote, 'too small for a satisfactory statistical comparison'.

Such comparisons of outcome before and after the introduction of new therapies were the most common method of evaluation. The possibility that the cases treated might have differed, or that some other change in management might have occurred at the same time, seems never to have been considered. But such 'trials' shed a fascinating light on the type of patient then commonly treated, and on the problems that doctors faced. Some sound all too familiar. In the same Lancet paper, Lister again:

At this period I was engaged in a perpetual conflict with the managing body who, anxious to provide hospital accommodation for the increasing population of Glasgow, for which the Infirmary was by no means adequate, were disposed to introduce additional beds beyond those contemplated in the original construction... so as to make me sometimes feel it a questionable privilege to be connected with the institution.

If we jump forward to 1937 we find a similar technique to that of Lister used to assess the effects of two sulphonamide compounds, Red Prontosil and sulphanilamide, in the treatment of streptococcal puerperal sepsis10. Again, the prevalence of the disease is remarkable. In Queen Charlotte's Hospital between 1932 and 1935 there were 82 cases of puerperal fever with blood culture positive for haemolytic streptococci, of whom 58 (72%) died. In 1936 and 1937, after the introduction of the sulphonamides, there were 22 cases with only 6 (27%) deaths. In this study positive blood cultures showed that at least reasonably similar types of patients were being treated, so we begin to move towards the basic requirements of modern clinical trials. The possibility that some other change in management was responsible for the improved results was considered in detail and the authors honestly admitted: 'it cannot be stated with certainty that the drugs in question have played an important part in the control of the infection... we are nevertheless justified in regarding that conclusion as highly probable'. The need for a reliable means of evaluating new treatments was evidently becoming apparent.

In the following year, 1938, the effect of M and B 693, sulphanilamide, on pneumococcal pneumonia was described11. The experimental method was an improvement on what had gone before in that patients given the new drug were compared with patients who were in hospital at the same time but who were untreated. The patients were carefully selected on the basis of their history and 'unequivocal physical signs of lobar consolidation', and X-ray confirmation was 'obtained in many'. However, by modern standards the way patients were selected for the new and for the old forms of treatment was inappropriate:

From March to the middle of June this year more than 200 cases of lobar pneumonia were admitted to Dudley Road Hospital, Birmingham. About half of these were treated with M and B 693, the others, admitted to the wards of our colleagues, serving as controls were receiving the usual routine non-specific treatment. Admission to each group occurred for the most part on alternate days. This paper records our findings in 100 of these cases, unselected except that three patients dying within 24 hours of admission have been excluded, since in these the drug did not have a fair trial. A corresponding selection was made in the control group.

The importance of the 'intention to treat' principle had evidently not yet been appreciated, and furthermore the trial was 'open' (as had been all the others I have described), with the physicians and the patients knowing what treatment was given. A variable dose of M and B 693 was used, and was progressively increased as the study proceeded. The paper described several individual patients in detail; one was 'veneected of 450 c.cm with prompt relief' (of what we are not told) and the effect of M and B 693 in causing methaemoglobinemia was studied. The results were dramatic: there were 8 deaths in 100 cases treated with M and B 693, compared with 27 in 100 controls. The latter fatality rate was shown to be similar to that observed in the previous year, indicating that the improved survival in the treated cases was probably due to the new drug.

If we now jump ahead 10 years, passing over the excitement of penicillin and the whole of the Second World War, we come to the study by Irving Wright of 'The evaluation of anticoagulants in the treatment of coronary thrombosis with myocardial infarction'12. This paper had a major impact on clinical practice.

The study included a crude - and as it turned out,
unsuccessful – attempt at establishing similar treatment and control groups of patients by a form of randomisation:

386 patients admitted to the participating services on even days received conventional treatment and constitute the control group. 432 patients admitted on odd days received anticoagulants in addition to conventional therapy and constitute the treated group.

The treated group was given heparin for 48 hours and then dicoumarol; the controls were not given either. The authors considered that the randomisation technique worked well, with treated and control groups containing respectively 77% and 76% of men, 24% and 22% of patients with previous myocardial infarction, and the average age was 60 and 59 years. However, the imbalance in numbers (treated 432, control 368) is suspicious of a bias in patient selection and one cannot help wondering if referring physicians deliberately weighted admissions on days when it was known that anticoagulants would be given. Be that as it may, 24% of the control patients died compared with 15% of those treated with anticoagulants and the authors concluded that ‘anticoagulation therapy should be used in all cases of coronary thrombosis with myocardial infarction unless a definite contraindication exists’. This was an important study, marred by its method of allocation to control groups. It led to several more studies of different anticoagulant regimens.

**The MRC trial of streptomycin in pulmonary tuberculosis**

At exactly the same time as Wright’s trial of anticoagulants in coronary thrombosis was published in *The American Heart Journal*, the *British Medical Journal* published the results of the Medical Research Council’s streptomycin trial. This trial, masterminded by Austin Bradford Hill (Fig 4), had almost all the characteristics of a modern clinical trial, and its publication marks a turning point in therapeutics.

For the first time, inclusion and exclusion criteria were laid down, even though these terms were not actually used. It was required that ‘all patients should have a similar type of disease’; they had to be aged 15–30 years, to have acute progressive bilateral pulmonary disease with bacteriological proof, they had to be unsuitable for ‘collapse therapy’ (artificial pneumothorax), and their chances of spontaneous regression had to be regarded as small yet there had to be some chance of improvement. Patients who fulfilled these criteria were – and this is the key point – allocated at random to treatment with bedrest plus streptomycin 2 g daily for 4 months, or to bedrest alone. Two different measures of success (what we would now call endpoints) were defined: death, and improvement on chest x-rays interpreted by a radiologist unaware of the treatment the patients had received. The only element of the design missing by modern standards was that the treatment itself was ‘open’, with doctors and patients being aware whether streptomycin was being given or not.

During the course of the trial the investigators noted for the first time something with which their successors would become all too familiar. The *British Medical Journal* paper records: ‘at first the impression was that cases of the type defined are seen often. In fact, such cases are not common.’ The number included was limited to 106 because ‘all the streptomycin available in the country was in any case being used, the rest of the supply being taken up for two rapidly fatal forms of disease, miliary and meningeal tuberculosis’.

Writing about the trial fifty years later, Bradford Hill, then 92, admitted that what is now sometimes called ‘type B statistics’ had been employed in defining trial size. ‘The new drug had been discovered in America in 1944 (and) in 1946 Britain literally had no currency. We had exhausted all our dollars in the War and our Treasury was adamant that we could only have a very small amount of streptomycin . . . we could only have enough to treat about 50 patients’. Bradford Hill doubted if the MRC would ever have agreed to his randomised and controlled trial if streptomycin had been freely available. However, he thought that ‘50 was probably enough to get a reliable answer . . . if streptomycin was really effective’.

And so it proved. Among the 52 patients in the control (bedrest only) group there were 14 deaths (27%) and among the 55 given streptomycin there were only 4 deaths (7%). The probability of this having occurred by chance was calculated to be less than 1 in 100. Moreover, improvements in the chest x-ray occurred in only 4 control patients (8%) compared with 28 (56%) in the streptomycin group.

The MRC streptomycin trial became the model for all its successors. Later trials demonstrated the clinical value of
many drugs – though few were as dramatically beneficial as streptomycin – but the MRC trial marked the end of important developments in clinical trial design. Thereafter the evaluation of new drugs followed a well-trodden path, and however important these new drugs might be, it is fair to say that the MRC streptomycin trial marked ‘the end of medical history’ in the same way that the development of liberal democracy marked ‘the end of history’.

**Double-blind trials**

The one key element missing in the streptomycin trial design was double-blind treatment. This was addressed in the next MRC trial, published in 1950 and again heavily influenced by Bradford Hill. This studied the effect of antihistamines in the prevention of the common cold: although it now seems a forlorn hope that such drugs might be useful, the study was based on a series of small clinical experiments that suggested that antihistamines were indeed helpful. Since the endpoints were less secure than in the streptomycin trial, in the antihistamine trial great care was taken to ensure that the blinding of treatment was maintained. The paper also shows that 40 years ago simple methods had been devised to ensure compliance with treatment:

Half the subjects received the drug under test, the other half received dummies indistinguishable from it and containing 1/4 gr. (16 mg) of phenobarbitone. It was felt that the latter would, in the dosage employed, have a mild sedative effect and so be acceptable as a control medication against which the efficacy of the antihistamine could be measured. Tablets were given at approximately 10 am and 7 pm, swallowing being supervised by the matron.

Apart from the double-blind design, the publication of the MRC antihistamine trial was important for two reasons: it showed that the results of poorly designed studies might be refuted by a single good trial, and it showed the importance of publishing ‘negative’ results.

**Clinical trials since 1950**

Since 1950 the most important development in clinical trial design has probably been the appreciation of the need for studies of an adequate size, coupled with a standardised way of expressing the validity of the trial results; otherwise, the ‘developments’ have ranged from disappointing to frankly misleading. I shall illustrate these points with examples of cardiovascular trials, which is easy to do because the evaluation of cardiovascular therapy has led the development of the whole science of clinical trials.

**Trial size and statistics**

The trials that I have described included relatively few patients, but they were successful because the treatments being studied – lemons for scurvy, streptomycin for tuberculosis – had dramatic effects. Since 1950 the norm has been the investigation of drugs with small, though real, benefits and it has been necessary to design trials capable of showing these small benefits.

The use of beta blockers in the secondary prevention of myocardial infarction makes the point. The first study, of propranolol, included 93 patients and found that the fatality of 35% in the placebo group was reduced to 16% among patients given active treatment. This was not a randomised trial and the results were not statistically significant. In a series of further relatively small trials, non-significant benefit from beta blockers was observed, but it was appreciated that to have a realistic chance of demonstrating a small benefit, a large number of patients had to be randomly allocated to treatment. If we just consider the trials in which beta blockers were given at least a couple of days after the infarction (the circumstances under which the benefit of beta blockade was eventually found most obvious), we find that practolol was the first drug to show clear benefit, in a trial which included just over 3,000 patients. However, practolol was withdrawn because of adverse effects and the first really successful trial of a safe beta blocker was that of timolol (1,884 patients).

The publication of numerous beta blocker trials with apparently conflicting results was associated with – and perhaps partly responsible for – new ways of expressing trial results. Instead of a measure of certainty – ‘the difference between treatment groups is highly significant’ (there was less than 1 chance in 100 that this was a random effect) – we have changed to a measure of uncertainty. We now express a result as a confidence interval, which denotes the limits within which the real trial result must lie. We also use a measure of the certainty of the two results lying within that interval – ‘the 95% confidence interval’ means that we can be 95% certain that the real result is within the stated range.

A final change in presentation of results has been the habit of presenting results as a risk reduction. Thus, if active treatment reduces an event rate from 50% to 25%, we talk about a 50% reduction of risk, but a reduction of 2% to 1% is also a 50% risk reduction, even though it is obviously of far less clinical importance. (For bibliography see Refs 17–19.)

The results of ‘late-entry’ beta blocker trials are shown in a standard format in Fig 5. Differences in trial results, and in our confidence in the result of each trial, are immediately apparent.

**Subgroup analysis**

Since there seem to have been no new drugs with comparable efficacy to oranges and lemons for scurvy, penicillin for staphylococcal infections, and streptomycin for pulmonary tuberculosis, clinical trials have in the last forty years often given equivocal results. An unfortunate habit has developed of ‘dredging’ trial results in an attempt to identify subgroups of patients, based on age, gender, disease
severity and concomitant treatment, who fared better than others in the active treatment group. The best proof that this was a grossly misleading approach came from a subset analysis of the results of the ISIS-2 trial of streptokinase in patients with acute myocardial infarction.20

This trial included a total of 17,187 patients; the fatality rate in the placebo group was 11.8% and in the streptokinase group 9.4%. In a subgroup of patients whose astrological birth sign was Gemini or Libra (a large sample, 2,799 patients) the death rate on placebo was 10.2%, while among those given streptokinase it was higher at 11.1%. All the benefit from streptokinase in the trial as a whole was due to patients with other birth signs. Presumably this is nonsense, and shows the danger of deriving results from subsets, even when they are large.

Surrogate endpoints

For reasons that are not always entirely clear, patients in clinical trials tend to have abnormally low event rates. The enthusiast might claim that this is the result of the extra care the trial patients tend to receive, but it seems more likely to be due to the inclusion in trials of low risk patients. Trial patients tend to be younger than average, and tend to have single diseases without complications. In trials where registers of all potential patients are kept, it is usually found that the inclusion rates are low, and that the highest event rates are among the excluded patients. Because of the need to obtain high event rates, and so give the intervention being tested a chance of demonstrating benefit, a habit developed of designing trials with 'surrogate endpoints'. This usually meant an alternative to death. Thus in a trial of an antiarrhythmic drug, the surrogate for death might be a reduction in arrhythmias, while in a heart failure trial the surrogate might be haemodynamic change such as an improvement in left ventricular ejection fraction.

Several trials have destroyed this concept. The CAST study showed that flecainide and encainide reduced arrhythmias detected by continuous ECG (Holter) monitoring, but their use (at least in patients with ischaemic heart
disease) was associated with a higher fatality rate. In trials of milrinone, enoximone, flosequinan and ibopamine, these drugs improved symptoms of heart failure but increased fatality. In thrombolysis, the GUSTO III trial showed that reteplase, the thrombolytic with the greatest coronary patency rate after myocardial infarction, was not associated with a low a fatality rate as was treatment with alteplase.

It is thus clear that 'surrogate' endpoints are hopelessly unreliable in drug evaluation.

Equivalence testing

The first drug of any completely new class has a reasonable chance of leading to a significant improvement in outcome, but later drugs in that class are unlikely to lead to further improvement of comparable magnitude. That is not to say that the later drugs are not worth having – they may well be easier to administer, have less severe adverse effects, and may be cheaper. This has led to the concept of 'equivalence' trials – trials that are designed to show that two drugs have an equivalent effect on the most important outcome. The problem with these trials is the definition of equivalence: this is not a statistical concept, but is simply a matter of what a clinician will accept.

The first true equivalence trial was INJECT, which aimed to show that a new thrombolytic, reteplase (which has the advantage of bolus administration), was equivalent to streptokinase in its effect on fatality in patients with acute myocardial infarction. The definition of equivalence used in INJECT was modest: the drugs would be considered equivalent if the fatality rate with reteplase could be no more than 1% higher than the rate seen with streptokinase. Since reteplase was expected to be superior to streptokinase, the confidence limits around the expected result could be quite wide and still permit equivalence to be claimed, and this allowed a modest trial size of 5,500 patients. The result was as expected: the fatality rate with reteplase was non-significantly better than that with streptokinase, but the 95% confidence intervals around that result excluded the possibility that the reteplase fatality rate was 1% higher than that of streptokinase. The drugs were thus – by the definition of the trial – equivalent.

This seemed to represent a small but useful advance in clinical trial design, but its limits were demonstrated only too clearly – and predictably – by the COBALT trial. Here the aim was to show that a double bolus of alteplase was an equivalent thrombolytic regimen to the less convenient continuous infusion: angiographic studies had suggested that a double bolus is actually more effective than an infusion. The definition of equivalence was that the fatality rate with the double bolus should not be greater than 0.4% more than the fatality rate with an infusion. Put statistically, the 95% confidence limits around the observed difference would exclude a maximum difference of 0.4%; 7,000 patients were included and the trial result was a fatality rate of 2.98% with double bolus compared with 7.53% with an infusion of alteplase; the double bolus proved slightly less effective and a claim for equivalence could not – in the terms of the trial – be claimed.

How many clinicians would worry about a difference between treatment group fatality rates of 0.43%? Surely the fact that investigators are even spending time thinking about this sort of difference shows how trivial are the 'advances' that some are now taking seriously.

Meta-analysis

I have saved until last the most disappointing development in therapeutics since the MRC trial of 1948 – the idea that clinical treatment can (indeed, should) be based on a summation, or meta-analysis, of the results of several trials. The idea is usually attributed to Cochrane, but James Lind should really be given the credit. In A treatise of the scurvy he wrote:

As it is no easy matter to root out prejudices, . . . it became requisite to exhibit a full and impartial view of what had hitherto been published on the scurvy, and that in a chronological order, by which the sources of these mistakes may be detected. Indeed, before this subject could be set in a clear and proper light, it was necessary to remove a great deal of rubbish.

Cochrane was appalled that there was no acceptance of the need to put together all available information, but it has to be remembered that Cochrane was working at a time when small trials were the norm, and since most therapeutic areas involved small benefits these trials were individually inconclusive. The early trials of beta blockers following myocardial infarction make the point. Acceptable ways of combining trial results were developed with enthusiasm by statisticians, but, in general, statisticians do not seem to have realised that it is futile to attempt to combine studies with different objectives, different patient types, and different doses of different drugs. This unthinking 'meta-analysis' is the equivalent of adding apples to oranges and talking about the effect of fruit – and though fruit worked for James Lind, that is not an excuse for meta-analysis.

There have been many misleading meta-analyses. For example, separate meta-analyses of the use of intravenous magnesium and of nitrates suggested that these drugs reduced mortality after myocardial infarction, but the definitive ISIS-4 study showed that neither was useful. Conversely, meta-analysing trials comparing the thrombolytics, streptokinase and alteplase, suggested that there was no difference between the two, but the studies included in the meta-analysis were performed before it was realised that the best results with alteplase were only obtained when it was given in a particular dosage regimen, and when it was combined with heparin. When a 'proper' trial was done – GUSTO – using alteplase in the most effective way, it was found that alteplase was superior to streptokinase.

A review of 19 meta-analyses which were followed by 12 definitive trials showed that without the definitive trial, accepting the meta-analysis result would have led to the
adoption of ineffective treatment in about one third of cases, the rejection of a useful treatment in another one third, and the adoption of a useful treatment in only about one third.

Clinical trials and evidence based medicine

It is now generally accepted that medical practice should be ‘evidence based’ – in other words, therapeutics should depend on the results of clinical trials. Unfortunately, the limitations of clinical trial methodology are frequently not appreciated. The need to establish, with statistical confidence, that a new remedy leads to a small benefit has, as I have shown, led to the introduction of new types of clinical trials in which we must have less clinical than statistical confidence.

Strictly speaking, the results of any drug trial relate only to the patients actually included, and to the individual drug and particular dose of that drug that was actually used. The problem is the wisdom of extrapolating the result of the trial to the individual patient who needs treating in ordinary clinical practice. Clinical trials often include patients within a certain age group and with certain clinical characteristics; in the real world patients are less well defined and do not fit the inclusion and exclusion criteria of the trial. In order to obtain a ‘clean’ answer, trials usually attempt to include only patients with a single disease; in practice patients frequently have multiple problems involving multiple therapy. Clinical trials, by their nature, tend to include low-risk patients, and when registers are maintained of all the patients who might have been considered for inclusion in the trial, it is clear that the trial’s results apply only to a small proportion of the generality of patients. Few direct comparisons within a class (beta blockers, ACE inhibitors, statins, for example) have been performed, and where they have been (as described above with thrombolytics), it is difficult to establish even at a philosophical level what we mean by ‘different’ and ‘equivalent’ What ‘evidence’ do we require to assume that all the drugs within a therapeutic class have a ‘class’ effect?

The concept of evidence-based medicine may have run its course, at least in its present form. All these qualified doubts, often expressed by anonymous expert committees, seem a far cry from the certainties of James Lind and Howard Florey. Uncertainty seldom makes history.

Conclusions

The MRC trial of streptomycin published in 1948 marked a turning point in clinical medicine. Untill that time major advances in therapeutics had been developed by experiments that would now be considered totally inadequate and unconvincing. Since then there have been many minor advances, developed on the basis of the methodology established with the streptomycin trial. That trial marked the end of major changes in clinical trial design and analysis: the methodological developments since have been small, and often more hindrance than help. However, clinical trials, inadequate though they often are, do provide the only avenue for therapeutic advance and the history of medical therapy is essentially the history of clinical trials. In the sense that Bradford Hill’s 1948 trial marked the onset of a universally accepted and essentially homogeneous way of investigating new treatments, we can look on the MRC streptomycin trial as ‘the end of medical history’.

References

1. Fukuyama F. The end of history and the last man. London: Penguin Books, 1992.
2. Abraham EP, Gardner D, Chain E, Heatley NG, et al. Further observations on penicillin. Lancet 1941;ii:177–89.
3. Elliott Smith A. Twenty years on. Oxford Medical School Gazette 1960:9–15.
4. Withering W. An account of the foxglove and some of its medical uses. Birmingham, Alabama: M Swinney, 1785. Reprinted: Birmingham, Alabama: The Classics of Medicine Library, Gryphon Editions Ltd, 1979.
5. The Digitalis Investigation Group. The effect of digoxin on mortality and morbidity in patients with heart failure. N Engl J Med 1997;336:525–33.
6. Louis PCA. Essay on clinical instruction, translated by P Martin. London: S Higheley, 1834: 26, 27, 28.
7. Lind J. A treatise of the scurvy. Printed by Sands, Murray, & Cochran for A Millar, in the Strand, London, 1753.
8. Lister J. On the effects of the antiseptic system of treatment upon the salubrity of a surgical hospital. Lancet 1870;i:4–6.
9. Lister J. On the effects of the antiseptic system of treatment upon the salubrity of a surgical hospital. Lancet 1870;ii:40–42.
10. Colebrook I, Purdie AW. Treatment of 106 cases of puerperal fever by sulphanilamide (streptocide). Lancet 1937;i:1237–42 and 1291–3.
11. Evans GM, Gaisford WF. Treatment of pneumonia with 2-(p-aminobenzensulphonamido) pyridine. Lancet 1938;i:14–19.
12. Wright IS, Marple CD, Beck DF. Report of the Committee for the evaluation of anti-coagulants in the treatment of coronary thrombosis with myocardial infarction. (A progress report on the statistical analysis of the first 800 cases studied by this committee). American Heart Journal 1948;36:801–15.
13. Medical Research Council Investigation. Streptomycin treatment of pulmonary tuberculosis. Br Med J 1948;ii:769–82.
14. Bradford-Hill A. Suspended judgment. Memories of the British Streptomycin Trial in Tuberculosis. The first randomized clinical trial. Control Clin Trials 1990;11:77–9.
15. Report by a special committee of The Medical Research Council. Clinical trials of anti-asthmatic drugs in the prevention and treatment of the common cold. Br Med J 1950;ii:426–31.
16. Snow PJD. Effect of propranolol in myocardial infarction. Lancet 1965;i:551–3.
17. Yusuf S, Peters R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomised trials. Prog Cardiovasc Dis 1985;27:335–71.
18. Australian and Swedish Pindolol Study Group. The effects of Pindolol on the two years mortality after complicated myocardial infarction. Eur Heart J 1983;4:367–75.
19. Boissel JP, Letzorovice A, Picolet H, Peyrieux JC. Efficacy of acebutolol after acute myocardial infarction (the APSI trial). The APSI investigators. Am J Cardiol 1990;66:251–60.
20. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17287 cases of suspected acute myocardial infarction: ISIS-2. Lancet 1988;i:439–49.
21. Wilcox RG, Hampton JR, Banks DC, Birkshead JS, et al. Trial of early nifedipine in acute myocardial infarction: the TRENT Study. Br Med J 1996;293:1204–8.
22. The Cardiac Arrhythmia Suppression Trial (CAST) 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.

23 Hampton JR, van Veldhuisen DJ, Kleber FX, Cowley AJ, et al. Randomised study of effect of ibopamine on survival in patients with advanced severe heart failure. *Lancet* 1997;349:971-7.

24 The Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO III) Investigators. *A comparison of reteplase with alteplase for acute myocardial infarction. N Engl J Med* 1997;337:1118-23.

25 International Joint Efficacy Comparison of Thrombolytics. Randomised, double-blind comparison of reteplase double-bolus administration with streptokinase in acute myocardial infarction (INJECT): trial to investigate equivalence. *Lancet* 1995;346:329-6.

26 The Continuous Infusion versus Double-Blind Administration of Alteplase (COBALT) Investigators. A comparison of continuous infusion of alteplase with double-bolus administration for acute myocardial infarction. *N Engl J Med* 1997;337:1124-30.

27 Cochrane AL. *A critical review with particular reference to the medical profession. In: Medicines for the year 2000. London: Office of Health Economics, 1979: 1-11.

28 ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. ISIS-4: A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet* 1995;345:669-85.

29 The GUSTO Investigators. An international randomised trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med* 1993;329:673-82.

30 LeLorier J, Gregoire G, Benhaddad A, Laperriere J, Derderian F. Discrepancies between meta-analyses and subsequent large randomised, controlled trials. *N Engl J Med* 1997;337:536-42.

31 Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn’t. *Br Med J* 1996;312:71-2.

Address for correspondence: Professor J R Hampton, Cardiovascular Medicine, University Hospital, Nottingham NG7 2UH.

---

**ROYAL COLLEGES OF PHYSICIANS**

**Notification of MRCP 1999 Examination Sittings**

The MRCP(UK) Examination will take place on the dates listed below. Candidates are requested not to submit applications for any sitting of the examination until the previous sitting has been completed.

### PART 1

| Sitting           | Date              | Closing Date         |
|-------------------|-------------------|----------------------|
| First Sitting     | 26th January 1999 | 27th November 1998   |
| Second Sitting    | 27th May 1999     | 26th March 1999      |
| Third Sitting     | 28th September 1999 | 30th July 1999      |

- Prospective candidates must have been qualified for at least 18 months and may enter through any of the Colleges listed below.
- Completed application forms accompanied by the necessary certificates and fee of £200 must reach the College of entry by the corresponding closing date listed above.

### PART 2

| Sitting           | Date              | Closing Date         |
|-------------------|-------------------|----------------------|
| First Sitting     | 12th January 1999 | 20th November 1998   |
| Second Sitting    | 4th May 1999      | 19th March 1999      |
| Third Sitting     | 7th September 1999 | 23rd July 1999      |

- Prospective candidates for Part 2 must complete a period of training of 2½ years from the date of graduation and must comply with the regulations concerning training in acute medicine.
- Completed application forms accompanied by the necessary documentation and fees must reach the College of entry by the corresponding closing date listed above. Early application is advised as places are limited and will be allocated on a ‘first come, first served’ basis. Examination fees: *Written Section* £200, *Oral and Clinical Section* £240. All three Colleges will require separate cheques.

Royal College of Physicians of Edinburgh
9 Queen Street, Edinburgh EH2 1JQ

Royal College of Physicians & Surgeons of Glasgow
242 St Vincent Street, Glasgow G2 5RJ

Royal College of Physicians of London
11 St Andrews Place, Regent’s Park, London NW1 4LE

Registered Charity No. 210508