

# The Development of Clinical Trials

SIMON DAY

Licensing Division, Medicines and Healthcare products Regulatory Agency, London, UK

## INTRODUCTION

This chapter traces some of the development of clinical trials – from very early anecdotal reports of informal evaluations of medicines (some not necessarily considered ‘medicines’ by today’s standards), medical practices, and so on – through to modern, well established principles, which include blinding, randomisation, clear protocols and analysis plans, etc. Some events have been milestones, whilst others have contributed in more modest ways to, what is now often considered as, the ‘gold standard’ of evidence for evaluating therapies.

## EARLIEST STORIES

The modern-day birth of clinical trials is usually considered to be the publication by the UK Medical Research Council in 1948 of a trial for the treatment of pulmonary tuberculosis with streptomycin, and we will return to this example later in the chapter. However, earlier but less well-documented examples do exist. The comparative concept of assessing therapeutic

efficacy has been known from ancient times. Lilienfeld<sup>1</sup> cites a description of a nutritional experiment involving a control group in the Book of Daniel from the Old Testament:

Then Daniel said to the guard whom the master of the eunuchs had put in charge of Hananiah, Mishael, Azariah and himself, ‘Submit to us this test for ten days. Give us only vegetables to eat and water to drink; then compare our looks with those of the young men who have lived on the food assigned by the king, and be guided in your treatment of us by what you see.’ The guard listened to what they said and tested them for ten days. At the end of ten days they looked healthier and were better nourished than all the young men who had lived on the food assigned them by the king. So the guard took away the assignment of food and the wine they were to drink, and gave them only the vegetables.

Daniel lived around the period 800 BC and although it may not be possible to confirm the accuracy of the account, what is clear is that when this passage was written – around 150 BC – the ideas certainly existed.

The passage from Daniel describes not just a control group, but a *concurrent* control group.

This fundamental element of comparative, experimental (and in this case clinical) research did not begin to be widely practised until the latter half of the twentieth century.

Much later than the book of Daniel, but still very early, is an example from the fourteenth century: it is a letter from Petrarch to Boccaceto cited by Witkosky:<sup>2</sup>

I solemnly affirm and believe, if a hundred or a thousand of men of the same age, same temperament and habits, together with the same surroundings, were attacked at the same time by the same disease, that if one followed the prescriptions of the doctors of the variety of those practicing at the present day, and that the other half took no medicine but relied on Nature's instincts, I have no doubt as to which half would escape.

During the fourteenth to sixteenth centuries, the Renaissance period was a time of great development in many forms ranging from art to science. This period provides other examples including an unplanned comparison of treatment of battlefield wounds. Packard<sup>3</sup> describes how, during a battle to capture the castle of Villaine in 1537, the surgeon Ambroise Paré was using the standard treatment (*sic*) of pouring boiling oil over soldiers' wounds. During the battle, he ran out of oil so he resorted to using a mixture of egg yolks, oil of roses and turpentine. The reason for this particular concoction seems unknown. The superiority of the new 'treatment' became evident the next day:

I raised myself very early to visit them, when beyond my hope I found those to whom I applied the digestive medicament feeling but little pain, their wounds neither swollen nor inflamed, and having slept through the night. The others to whom I had applied the boiling oil were feverish with much pain and swelling about their wounds. Then I determined never again to burn thus so cruelly by arquebusses.

Today, we might (at best) call such an experience a 'natural experiment'; at worst we would simply consider it an anecdotal experience, completely

confounded with time and so possibly also type and severity of wounds, weather conditions and a host of other unknown factors.

Perhaps the most famous historical example of a planned, prospective controlled, comparative, clinical trial is from the eighteenth century: that where Lind<sup>4</sup> found oranges and lemons to be the most effective of six dietary treatments for scurvy on board ships. His account (reproduced from Anderson)<sup>5</sup> reads thus:

On 20th of May, 1747, I took 12 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 their knees. They laid together in one place being the proper treatment for the sick in the forehold; and had one diet common to all, water gruel, sweetened with sugar in the morning; fresh mutton broth often for dinner; at other times pudding, boiled biscuit with sugar and for supper, barley and raisins, rice and currants, sago and wine, or the like. Two of these were ordered each a quart of cider a day. Two others took 25 drops of elixir vitriol three times a day upon an empty stomach; using a gargle strongly acidulated with it for their mouths. Two others took two spoonfuls of vinegar three times a day upon an empty stomach, having the gruel and other food well acidulated with it, and also the gargle for their mouths. Two of the worse patients were put on a course of sea water. Of this they drank half a pint a day and sometimes more or less as it operated by way of gentle physic. Two others had each two oranges and one lemon given them every day. These they eat with greediness at different times, upon an empty stomach. They continued for six days under this course, having consumed the quantity that could be spared. The two remaining patients took the bigness of a nutmeg three times a day of an electuary recommended by a hospital surgeon, made of garlic, mustard seed, *Rad. Raphan.*, balsam of Peru, and gum Myrrh, using for drink barley water well acidulated with tamarinds; by a concoction of which with the addition of cream of tartar they were greatly purged three or four times during the course.

The consequence was that the most sudden and 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 spots were not indeed at that time quite off his body, nor his gums sound; but without any other medicine, than a gargarism of elixir vitriol, he became quite healthy before we came into Plymouth, which was on the 16th June. The other was the best recovered of any in his condition, and being now deemed pretty well, was appointed nurse to the rest of the sick.

Pierre-Charles-Alexandre Louis,<sup>6</sup> a nineteenth-century clinician and pathologist, introduced the numerical aspect to comparing treatments. His idea was to compare the results of treatments on groups of patients with similar degrees of disease (which is not quite the case with Lind), and so to truly compare ‘like with like’:

I come now to therapeutics, and suppose that you have some doubt as to the efficacy of a particular remedy: How are you to proceed? . . . You would take as many cases as possible, of as similar a description as you could find, and would count how many recovered under one mode of treatment, and how many under another; in how short a time they did so, and if the cases were in all respects alike, except in the treatment, you would have some confidence in your conclusions; and if you were fortunate enough to have a sufficient number of facts from which to deduce any general law, it would lead to your employment in practice of the method which you had seen oftenest successful.

‘Like with like’ was an important step forward from Lind’s investigation of the treatment of scurvy. Note, although early in Lind’s passage he says that ‘Their cases were as similar as I could have them’, later he acknowledges (partly through a clear and detailed description of the study) that the two worst cases both received the same treatment: ‘Two of the worst patients were put on a course of sea water.’ His use of the verb ‘put’ (rather than, perhaps, ‘received’) implies an intention on Lind’s part. Perhaps he expected that the sea water might be the best treatment. It was more than a century later when Bradford Hill used a formal randomisation procedure for creating groups of cases that were ‘in all respects alike, except in the treatment’.

## RANDOMISATION

The use of randomisation was a major contribution to experimental design, put forward by the statistician and geneticist R.A. Fisher in agricultural trials (see, for example, Fisher,<sup>7</sup> Fisher and McKenzie).<sup>8</sup> Fisher randomised plots of crops to receive different treatments. In clinical trials, there had been early schemes to use ‘group randomisation’ whereby patients were divided into two groups and then the treatment for each group was randomly assigned. The Belgian medicinal chemist van Helmont<sup>9</sup> described an early example of this:

Let us take out of the hospitals, out of the Camps, or from elsewhere, 200, or 500 poor People that have Fevers, Pleurisies, &c, Let us divide them into halves, let us cast lots, that one half of them may fall to my share, and the others to yours . . . we shall see how many funerals both of us shall have: *But* let the reward of the contention or wager, be 300 florens, deposited on both sides.

Considering modern-day standards of trials it is interesting to compare and contrast features such as:

- a description of some sort of inclusion criteria;
- a pre-specified, clinically relevant, endpoint (although today we might use the more politically correct term ‘all-cause mortality’); and
- some indication of sample size (although not very definitively chosen).

More recently, Amberson and McMahon<sup>10</sup> used group randomisation in a trial of sanocrysin for the treatment of pulmonary tuberculosis. Today, the more common term to describe such trials is ‘cluster’ randomised trials; a good review is contained in an issue of the review journal *Statistical Methods in Medical Research* (see Donner and Klar).<sup>11</sup>

Systematic assignment was used by Fibiger,<sup>12</sup> who alternately assigned diphtheria patients to serum treatment or an untreated control group. Alternate assignment is frowned upon today,

partly because knowledge of the future treatment allocations may selectively bias the admission of patients into the treatment group,<sup>13</sup> also because any unknown patterns of patient presentation may turn out to be correlated with the treatment assignment. ‘Proper’ randomisation<sup>14</sup> will avoid this possibility. Diehl *et al.*<sup>15</sup> reported a common cold vaccine study with University of Minnesota students as subjects where proper random assignment and blinding of patients to treatments appears to have been used:

At the beginning of each year...students were assigned at random...to a control group or an experimental group. The students in the control groups...received placebos...All students thought they were receiving vaccines...Even the physicians who saw the students...had no information as to which group they represented.

However, Gail<sup>16</sup> points out that although this appears to be a randomised clinical trial, a further unpublished report by Diehl clarifies that this is another instance of systematic assignment:

At the beginning of the study, students who volunteered to take these treatments were assigned alternately and without selection to control groups and experimental groups.

Bradford Hill, in the study of streptomycin in pulmonary tuberculosis,<sup>17</sup> used random sampling numbers in assigning treatments to subjects, so that the subject was the unit of randomisation. This study is now generally acknowledged to be the ‘first properly randomised clinical trial’ – although it was not fully blinded, as discussed below.

Later, Bradford Hill and the British Medical Research Council continued with further randomised trials: chemotherapy of pulmonary tuberculosis in young adults,<sup>18</sup> antihistaminic drugs in the prevention and treatment of the common cold,<sup>19</sup> cortisone and aspirin in the treatment of early cases of rheumatoid arthritis,<sup>20,21</sup> and long-term anticoagulant therapy in cerebrovascular disease.<sup>22</sup>

## BLINDING

The common cold vaccine study published by Diehl *et al.*<sup>15</sup> cited earlier, in which University of Minnesota students were alternately assigned to vaccine or placebo, was a masked (or blinded) clinical trial:

All students thought they were receiving vaccines...Even the physicians who saw the students...had no information as to which group they represented.

Partial blinding was used in the early Medical Research Council trials in which Bradford Hill was involved. Thus, in the first of those trials, the study of streptomycin in tuberculosis,<sup>17</sup> although patients and their treating physicians were not blinded to the treatment assignment, the X-ray films were viewed by two radiologists and a clinician, each reading the films independently and not knowing if the films were of C (control, bed-rest alone) or S (streptomycin and bed-rest) cases.

Bradford Hill<sup>23</sup> noted in respect of using such blinding and randomisation:

If [the clinical assessment of the patient’s progress and of the severity of the illness] is to be used effectively, without fear and without reproach, the judgements must be made without any possibility of bias, without any overcompensation for any possible bias, and without any possible accusation of bias.

Simply overcoming bias may not be sufficient: overcoming any possible *accusation of bias* is an important justification for blinding and randomisation. It is not clear if Bradford Hill considered the blind assessment of the X-rays (hence, the outcome measure) was adequate, or whether blinding of patients and treating physicians was necessary. Today, blinding (including treatment allocation concealment) and randomisation are considered the two most important (although not necessarily completely adequate) aspects of a good, well-controlled clinical trial.

In the second MRC trial, the antihistamine common cold study,<sup>19</sup> placebos, indistinguishable from the drug under test, were used. Here, Bradford Hill noted:

in [this] trial . . . feelings may well run high . . . either of the recipient of the drug or the clinical observer, or indeed of both. If either were allowed to know the treatment that had been given, I believe that few of us would without qualms accept that the drug was of value – if such a result came out of the trial.

In the United States, the National Institutes of Health started their first randomised trial in 1951. It was a National Heart Institute study of adrenocorticotrophic hormone (ACTH), cortisone and aspirin in the treatment of rheumatic heart disease.<sup>24</sup> This was followed in 1954 by a randomised trial of retrolental fibroplasia (now known as retinopathy of prematurity), sponsored by the National Institute of Neurological Diseases and Blindness.<sup>25</sup> During the four decades following the pioneering trials of the 1940s and 1950s, there was a large growth in the number of randomised trials not only in Britain and the United States, but also in Canada and mainland Europe.

## ETHICS

Experimentation in medicine is as old as medicine itself and there is nothing necessarily wrong with that. Some experiments on humans have, however, been conducted without concern for the welfare of the subjects, who may have been prisoners or disadvantaged people. Katz<sup>26</sup> provides examples of nineteenth-century studies in Russia and Ireland of the consequences of intentionally infecting people with syphilis and gonorrhoea. McNeill<sup>27</sup> describes how, during the same period in the United States, physicians put slaves into pit ovens to study heat stroke, and poured scalding water over them as an experimental cure for typhoid fever. He even describes how one slave had two fingers amputated in a ‘controlled trial’, one finger with anaesthetic and one without! The benefits of the strength of causal evidence

obtained from a well-controlled trial hardly outweigh the ethical unacceptability.

Unethical experiments on human beings have continued into the twentieth century and have been described by, for example, Beecher,<sup>28</sup> Freedman<sup>29</sup> and McNeil.<sup>27</sup> In 1932 the US Public Health Service began a study in Tuskegee, Alabama, of the natural progression of untreated syphilis in 400 black men. The intentional withholding of treatment may be the first point of unacceptable ethics; the fact that the experiment was restricted to black men only (while whites received treatment) is yet further concern. It is quite remarkable that the study continued right up until 1972 when a newspaper reported that the subjects were uninformed or misinformed about the purpose of the study.<sup>29</sup> Shirer,<sup>30</sup> amongst others, describes how during the Nazi regime from 1933 to 1945, German doctors conducted experiments, mainly on Jews, but also on Gypsies, mentally disabled persons, Russian prisoners of war and Polish concentration camp inmates. The Nazi doctors were later tried and found guilty of these atrocities in 1946–7 at Nuremberg and this led to the writing, by three of the trial judges, of the Nuremberg Code (see US Government Printing Office).<sup>31</sup> This was the first international effort to lay down ethical principles of clinical research. Principle 1 of the Code states:

The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, overreaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision.

Other principles of the Code are that experiments should yield results for the good of society, that unnecessary suffering and injury should be avoided, and that the subject should be free to withdraw from the experiment at any time

and for any reason. Modern-day standards of acceptable ethics go beyond simply obtaining the patient's consent: a study that is unethical from the physician's (or society's) point of view cannot be considered acceptable simply by a patient giving his or her consent.

Other early advocates of informed consent were Charles Francis Withington and William Osler. Withington<sup>32</sup> realised the 'possible conflict between the interests of medical science and those of the individual patient', and concluded in favour of 'the latter's indefensible rights'. Osler<sup>33</sup> insisted on informed consent in medical experiments. Despite this early advocacy, and the 1946–7 Nuremberg Code, the application of informed consent to medical experiments did not take hold until the 1960s. Bradford Hill,<sup>34</sup> based on his experience in a number of early randomised clinical trials sponsored by the MRC, believed that it was not feasible to draw up a detailed code of ethics for clinical trials that would cover the variety of ethical issues that came up in these studies. He also considered that the patient's consent was not warranted in all clinical trials – a view that would not be generally supported today. Gradually the medical community has come to recognise the need to protect the reputation and integrity of medical research (as well as protecting patients and research subjects) and in 1955 a human experimentation code was adopted by the Public Health Council in the Netherlands.<sup>35</sup> Later, in 1964, the World Medical Assembly issued the Declaration of Helsinki<sup>36</sup> essentially adopting the ethical principles of the Nuremberg Code, with consent being 'a central requirement of ethical research' (see Faden *et al.*).<sup>37</sup> The Declaration of Helsinki has been updated and amended several times: Tokyo 1975, Venice 1983, Hong Kong 1989, Cape Town 1996 and Edinburgh 2000.

## THE PHARMACEUTICAL INDUSTRY

The impact of advances in clinical trial thinking has had a major impact on the pharmaceutical industry, which relies heavily on trials

for providing evidence of efficacy and safety of the products that it develops. However, in return, the industry has had a major influence on standards of clinical trials for exactly similar reasons – namely that it carries out so many of them.

Until around about the time of the thalidomide disaster (see, for example, Shah),<sup>38</sup> new medicines were licensed largely upon evidence that they were safe. Efficacy was less of an issue. Changes were introduced following thalidomide (although it should be noted that this was not the only product that prompted changes). It is interesting to observe that in Britain, the body that advises the Licensing Authority has, until recently, been called the Committee on Safety of Medicines (see, for example, Day).<sup>39</sup> In October 2005, along with other changes to the procedural aspects of licensing medicines (although not the scientific aspects), a new body, the Commission on Human Medicines, was established. It has a similar remit to the former Committee, although its change of name – clearly to encompass more than only safety – is a better descriptor.

Setting aside the semantics of the naming of advisory committees, licensing of new medicines today requires (amongst other things) clear and convincing evidence of efficacy. This, of course, best comes from high-quality clinical trials. Numerous guidelines (to guide industry as well to set common standards for assessment) have been developed covering many aspects of drug development and the demonstration of safety and efficacy in clinical trials. Amongst these – and being the most overarching – are those of the International Conference on Harmonisation ([www.ich.org](http://www.ich.org)) and of the guidelines produced by that body, 'E9' covers statistical principles for clinical trials.<sup>40,41</sup> That document has served as an excellent state of the art for most aspects of clinical trial design, conduct, analysis and reporting. However, with ever-increasing commercial pressures to bring new products to the marketplace more quickly, statisticians and other scientists working in the pharmaceutical industry have a keen interest to use – and often contribute to – new developments in clinical trial design and analysis.

## DATA MONITORING

In the modern randomised clinical trial, particularly for trials of life-threatening conditions, the accumulating data are often monitored for safety and efficacy by an independent data monitoring committee (see, for example, Ellenberg *et al.*)<sup>42</sup> One of the earliest examples of this was in 1968 when such a committee was established to serve the Coronary Drug Project, a large multicentre trial sponsored in the United States by the National Heart Institute of the National Institutes of Health.<sup>43,44</sup> In 1967, after a presentation of interim outcome data by the study coordinators to all participating investigators, Thomas Chalmers, clearly with great insight, wrote to the policy board chairman expressing his concern:

that knowledge by the investigators of early nonstatistically significant trends in mortality, morbidity, or incidence of side effects might result in some investigators – desirous of treating their patients in the best possible manner, i.e., with the drug that is ahead – pulling out of the study or unblinding the treatment groups prematurely.

We can note here the distinction between collective ethics and individual ethics – what is best for the trial, as opposed to what might be best for the individual patients. Following this letter, a more formal data and safety monitoring committee was established for the Coronary Drug Project consisting of scientists who were not contributing data to the study. Thereafter, the practice of sharing accumulating outcome data with the study's investigators, and others closely connected with the study, was discontinued. The data safety and monitoring committee assumed responsibility for deciding when the accumulating data warranted changing the study protocol or terminating the study.

The first formal recognition of the need for interim analyses, and the recognition that such analyses affect the probability of the Type I error, came with the publication in the 1950s of papers on sequential clinical trials by Bross<sup>45</sup> and Armitage.<sup>46</sup> The principal advantage of a sequential trial over a fixed sample size trial is

that, when the length of time needed to reach an endpoint is short, e.g. weeks or months, the sample size required to detect a substantial benefit from one of the treatments is reduced from what it would be in a more traditional 'fixed sample size' design.

In the 1970s and 1980s solutions to interim analysis problems came about in the form of group sequential methods and stochastic curtailment.<sup>47–49</sup> In the group sequential trial, the frequency of interim analyses is usually limited to a small number, say between three and six. The boundaries proposed by Pocock<sup>50</sup> use constant nominal significance levels; those proposed by Haybittle<sup>51</sup> and Peto<sup>52</sup> use stringent significance levels for all except the final test; in the O'Brien–Fleming<sup>53</sup> method, stringency gradually decreases; in the method by Lan and DeMets,<sup>54</sup> the total Type I error probability is gradually spent in a manner that does not require the timing of analyses to be prespecified. More details of these newer methods in the development of clinical trials are given in the next chapter.

## RECENT YEARS . . .

In recent years we have seen a huge increase in the number of trials carried out and published, and in the advancement of methodological aspects relating to trials. Whilst many see the birth of clinical trials (certainly in their modern-day guise) as being the MRC streptomycin trial,<sup>17</sup> there remains some controversy (see, for example, D'Arcy Hart,<sup>55,56</sup> Gill<sup>57</sup> and Clarke<sup>58</sup>). However, it is interesting to note that one of the most substantial reviews of historical aspects of trials is based on Bull's work for a 1951 MD thesis.<sup>59</sup> He cites 135 historical examples and other supporting references – but no mention of Bradford Hill and the MRC. The modern-day story of clinical trials perhaps begins where Bull ended.

Today, there are many academic papers devoted to the methodology of clinical trials; there are many books on the general methods of trials, as well as others on specific technical points

of trials and those in specific therapeutic areas. There are journals specifically devoted to clinical trials (*Clinical Trials: Journal of the Society for Clinical Trials* and *Contemporary Clinical Trials*) and there is a professional society – the Society for Clinical Trials ([www.sctweb.org](http://www.sctweb.org)).

## ACKNOWLEDGEMENTS

An earlier version of this chapter was based on work by F. Ederer, in: Armitage P and Colton T (Eds) *Encyclopedia of Biostatistics*, 1998: John Wiley & Sons, Ltd, Chichester.

## REFERENCES

- Lilienfeld AM. *Ceteris paribus*: the evolution of the clinical trial. *Bull Hist Med* (1982) **56**: 1–18.
- Witkosky SJ. *The Evil That Has Been Said of Doctors: Extracts From Early Writers*, trans with annotations by TC Minor, The Cincinnati Lancet-Clinic, Vol. 41, New Series Vol. 22, 447–448 (1889).
- Packard FR. *The Life and Times of Ambroise Paré*, 2nd ed. New York: Paul B. Hoeber (1921) 27, 163.
- Lind J. *A Treatise of the Scurvy*. Edinburgh: Sands Murray Cochran, (1753) 191–3.
- Anderson M. *An Introduction to Epidemiology*. London: Macmillan, (1980) 163–4.
- Louis PCA. The applicability of statistics to the practice of medicine. *London Med Gaz* (1837) **20**: 488–491.
- Fisher RA. The arrangement of field experiments. *J Min Agric* (1926) **33**: 503–13.
- Fisher RA, and McKenzie WA, Studies in crop variation: II. The manurial response of different potato varieties. *J Agric Sci* (1923) **13**: 315.
- van Helmont JB. *Oriatrike or Physik Refined*, trans J Chandler. London: Lodowick Loyd (1662).
- Amberson B, McMahan PM. A clinical trial of sanocrysin in pulmonary tuberculosis. *Am Rev Tuberc* (1931) **24**: 401.
- Donner A, Klar N. Cluster randomization trials (editorial). *Stat Methods Med Res* (2000) **9**: 79–80.
- Fibiger I. Om Serum Behandlung of Difteri. *Hospitalstidende* (1898) **6**: 309–25, 337–50.
- Altman DG, Bland JM. Treatment allocation in controlled trials: why randomise? *Br Med J* (1999) **318**: 1209.
- Day S. Randomisation. In: Everitt B and Palmer C. *Encyclopaedic Companion to Medical Statistics*, London: Edward Arnold, (2005) 291–4.
- Diehl HS, Baker AB, Cowan DW. Cold vaccines: an evaluation based on a controlled study. *J Am Med Assoc* (1938) **111**: 1168–73.
- Gail MH. Statistics in Action. *J Am Stat Assoc* (1996) **91**: 1–13.
- Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *Br Med J* (1948) **ii**: 769–82.
- Medical Research Council. Chemotherapy of pulmonary tuberculosis in young adults. *Br Med J* (1952) **i**: 1162–8.
- Medical Research Council. Clinical trials of anti-histaminic drugs in the prevention and treatment of the common cold. *Br Med J* (1950) **ii**: 425–31.
- Medical Research Council. A comparison of cortisone and aspirin in the treatment of early cases of rheumatoid arthritis – I. *Br Med J* (1954) **i**: 1223–7.
- Medical Research Council. A comparison of cortisone and aspirin in the treatment of early cases of rheumatoid arthritis – II. *Br Med J* (1955) **ii**: 695–700.
- Hill AB, Marshall J, Shaw DA. A controlled clinical trial of long-term anticoagulant therapy in cerebrovascular disease. *Q J Med* (1960) **29** (NS): 597–608.
- Hill AB. *Statistical Methods in Clinical and Preventative Medicine*. Edinburgh: E and S Livingstone (1962).
- Rheumatic Fever Working Party. The evolution of rheumatic heart disease in children: five-year report of a co-operative clinical trial of ACTH, cortisone, and aspirin. *Circulation* (1960) **22**: 505–15.
- Kinsey VE. Retrolental fibroplasia. *Am Med Assoc Arch Ophthalmol* (1956) **56**: 481–543.
- Katz J. The Nuremberg Code and the Nuremberg Trial. A reappraisal. *J Am Med Assoc* (1996) **276**: 1662–6.
- McNeill PM. *The Ethics and Politics of Human Experimentation*. Cambridge: Press Syndicate of the University of Cambridge (1993).
- Beecher HK. Ethics and clinical research. *New Engl J Med* (1966) **274**: 1354–1360.
- Freedman B. Research, unethical. In: Reich WT, ed, *Encyclopedia of Bioethics*. New York: Free Press, (1995) 2258–61.
- Shirer WL. *The Rise and Fall of the Third Reich*. New York: Simon and Schuster (1960).
- US Government Printing Office. *Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10*, Vol. 2. Washington DC: US Government Printing Office (1949) 181–2.



32. Withington CF. *Time Relation of Hospitals to Medical Education*. Boston: Cupples Uphman (1886).
33. Osler W. The evolution of the idea of experiment. *Trans Congr Am Physicians Surg* (1907) **7**: 1–8.
34. Hill AB. Medical ethics and controlled trials. *Br Med J* (1963) **i**: 1043–9.
35. Netherlands Minister of Social Affairs and Health. *4 World Med J* (1957) 299–300.
36. World Medical Assembly. *Declaration of Helsinki: recommendations guiding physicians in biomedical research involving human subjects*. Helsinki: World Medical Assembly (1964).
37. Faden RR, Beauchamp T, King NMP. *A History of Informed Consent*, New York: Oxford University Press, (1986).
38. Shah RR. Thalidomide, drug safety and early drug regulation in the UK. *Adverse Drug React Toxicol Rev* (2001) **20**: 199–255.
39. Day S. Medicines and Healthcare products Regulatory Agency (MHRA). In: Everitt B and Palmer C. *Encyclopaedic Companion to Medical Statistics*. London: Edward Arnold, (2005) 218–219.
40. Lewis JA. Statistical Principles for Clinical Trials (ICH E9): an introductory note on an international guideline. *Stat Med* (1999) **18**: 1903–4.
41. ICH E9 Expert Working Group. Statistical principles for clinical trials: ICH harmonised tripartite guideline. *Stat Med* (1999) **18**: 1905–42.
42. Ellenberg SS, Fleming TR, DeMets DL. *Data Monitoring Committees in Clinical Trials*. Chichester: John Wiley & Sons Ltd (2002).
43. Canner P. Monitoring of the data for adverse or beneficial treatment effects. *Controlled Clin Trials* (1983) **4**: 467–83.
44. Friedman L. The NHLBI model: a 25-year history. *Stat Med* (1993) **12**: 425–31.
45. Bross I. Sequential medical plans, *Biometrics* (1952) **8**: 188–295.
46. Armitage P. Sequential tests in prophylactic and therapeutic trials. *Q J Med* (1954) **23**: 255–74.
47. Armitage P. Interim analysis in clinical trials, *Stat Med* (1991) **10**: 925–37.
48. Friedman LM, Furberg CD, and DeMets DL. *Fundamentals of Clinical Trials*, 2nd edn. Boston: Wright (1985).
49. Pocock SJ. *Clinical Trials: A Practical Approach*. Chichester: John Wiley & Sons, Ltd (1983).
50. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* (1977) **64**: 191–9.
51. Haybittle JL. Repeated assessment of results of cancer treatment. *Br J Radiol* (1971) **44**: 793–7.
52. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith P. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. 1. Introduction and design. *Bri J Cancer* (1976) **34**: 585–612.
53. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* (1979) **35**: 549–56.
54. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika* (1983) **70**: 659–63.
55. D'Arcy Hart P. History of randomised controlled trials (letter to the Editor). *Lancet* (1972) **i**: 965.
56. D'Arcy Hart P. Early controlled clinical trials (letter to the Editor). *Br Med J* (1996) **312**: 378–9.
57. Gill DBEC. Early controlled trials (letter to the Editor). *Br Med J* (1996) **312**: 1298.
58. Clarke M. Early controlled trials (letter to the Editor). *Br Med J* (1996) **312**: 1298.
59. Bull JP. The historical development of clinical therapeutic trials. *J Chron Dis* (1959) **10**: 218–248.

