

CHAPTER 1

The Rationale for Treatment

A Brief History

The development of modern medicine has rightly been described as ‘the greatest benefit to mankind’ [1]. Vaccination, anaesthetics, aseptic surgery, antibiotics, insulin for diabetes and drugs to prevent and treat heart disease form part of a very long list of treatments that have transformed healthcare. These are the fruits of many years of careful clinical investigation supported by extensive research. However, medicine has a checkered history in which ineffective treatments were widely used: as Benjamin Franklin pithily remarked in the eighteenth century, ‘God heals, and the doctor takes the fees’ [2]. Studies of the history of medicine show that the basis on which treatments have been used has varied greatly across the centuries [3]. This chapter explores the rationale behind the use of treatments, and the way this

Evidence in Medicine: The Common Flaws, Why They Occur and How to Prevent Them,
First Edition. Iain K Crombie.

© 2021 John Wiley & Sons Ltd. Published 2021 by John Wiley & Sons Ltd.

has changed over time. It concludes by describing the progress towards present-day methods for testing the effectiveness of treatments.

THEORY AS JUSTIFICATION FOR TREATMENT

In ancient times diseases were attributed to supernatural causes, spirits and demons. Treatments involved spells and prayers, or the wearing of amulets, which were intended to drive the malign forces from the patient [4]. Theories gradually evolved towards biological and physical causes of disease, with treatments involving minor surgery and drugs (usually based on plant extracts, minerals and metals). In Western medicine, one of the most influential of these theories was the doctrine of the four humours. It held that good health was enjoyed when four humours (the fluids: blood, phlegm, black bile and yellow bile) were in balance, with an excess of one humour causing disease [5]. Treatment for illness focused on restoring the balance by removing some of the excess humour from the body. This could be achieved by bloodletting (cutting open a vein or by applying a leech), or by losing fluid with a purgative or blistering the skin. This treatment was almost always harmful, although it often appeared to give short-term relief of the symptoms of acute inflammations [6]. The most notable casualty of bloodletting was George Washington, first president of the United States. He was suffering from a serious upper respiratory tract infection, for which his doctors extracted approximately 2.4L of blood over about 12 hours. He died 33 hours later, probably from the combination of the infection and the treatment given [5]. When the practice of bloodletting was challenged in the nineteenth century, a leading physician, William Stokes, commented that it was hard to believe ‘that the fathers of British medicine were always in error, and that they were bad observers and mistaken practitioners’ [7]. This cautionary tale of bloodletting suggests that theory and clinical experience may be unreliable guides to the effectiveness of a treatment.

A more recent but widely (mis)used theory was that bed rest was beneficial for a variety of ailments. Its popularity has been traced to a series of lectures in the middle nineteenth century by

John Hilton, president of the Royal College of Surgeons [8, 9]. Initially recommended for recovery following orthopaedic procedures [10], it was soon used for conditions including myocardial infarction, pulmonary tuberculosis, rheumatic fever and psychiatric illnesses [9]. Bed rest was particularly popular in pregnancy, where it was recommended for complications such as threatened abortion, hypertension or preterm labour [11]. The theory was that if rest helped to mend broken bones, then it would also heal other organs [9]. The benefits of bed rest were thought to include reduced demands on the heart, conservation of metabolic resources for healing and avoidance of stress [12]. Its use began to be challenged in the middle of the twentieth century, as evidence grew on the adverse effects of bed rest; it is now known to cause impairment of cardiovascular, haematological, musculoskeletal, immune and psychological functions [9, 12]. Bed rest is an example of a treatment based on beliefs about benefit that endured in the face of substantial evidence of harm [8, 11].

TESTING ON A SERIES OF PATIENTS

The transition, from treatments based on theory to the use of evidence derived from empirical studies, was a gradual process. A simple, and common, method was to give a treatment on a series of patients, then observe its impact on disease. A good example is the use of the leaves of the willow tree for inflamed joints, a treatment dating back to the ancient Egyptians [13]. Clinical observation confirmed the benefits: application of a decoction of willow leaves to inflamed skin reduced the swelling. Extracts of willow leaves and bark were also used for fever and pain by the Greeks from the fifth century BCE [14]. An important step in the use of the willow was taken by the Reverend Edward Stone in 1763. He administered a solution of powdered willow bark to 50 patients with fever, judging the treatment a great success [14, 15]. The active ingredient of the willow, salicin, was isolated in the 1820s [13, 15]. This drug was tested by a Dundee physician, T.J. MacLagan, who administered it to a series of patients with acute rheumatism. Not only was the

treatment successful, it demonstrated antipyretic, analgesic and anti-inflammatory effects [15]. Salicin was recognised to be an important drug, but its long-term use was limited because gastric irritation, nausea and vomiting were common side effects. The pharmaceutical arm of the Bayer company searched for a safer alternative, and successfully modified salicin to produce a new chemical with fewer side effects [13, 15]. That drug, aspirin, is now the most widely used medicine in the world [14].

Another example of evidence from a series of patients is the discovery of insulin for the treatment of diabetes. This was undoubtedly ‘one of the most dramatic events in the history of the treatment of disease’ [16]. Research, in the late nineteenth century, had shown that removal of an animal’s pancreas ‘produced severe and fatal diabetes’ [17]. Over the following 30 years many researchers tried to isolate a pancreatic extract that could control blood sugar levels. They had little success, as the extracts had only a transitory effect on blood sugar and caused unacceptable side effects (vomiting, fever and convulsions) [18, 19]. In October 1920 Frederick Banting, a young Canadian doctor, was preparing a lecture on the pancreas [16]. The research he was reading led him to think that the active ingredient was being destroyed by the digestive enzymes in the pancreas, and that this could be prevented by ligating the pancreatic ducts. Banting began the experiments with extracts of the ligated pancreas in May 1921 [17]. By January 1922 a purified extract had been obtained. This proved successful in treating a 14-year-old boy, and in February a further six patients were treated with equally favourable results [16]. The discovery was announced in April to international acclaim; the Nobel prize was awarded to Banting, and one of his colleagues, Dr Macleod, in 1923 [16].

COMPARING GROUPS

Case series can provide support for a treatment if, as with insulin, the benefits are immediate and substantial. But observations on a set of patients are often not sufficient to identify whether a treatment is truly effective. Consider the management of gunshot wounds in

the sixteenth century. At that time it was believed that the bullet introduced poison into the body, and that cauterising the wound with boiling oil mixed with treacle would detoxify it [20, 21]. The treatment was very unpleasant, but was thought to save lives. Force of circumstances led the French barber-surgeon, Ambroise Paré, to use a different treatment. During the Italian war of 1536–1538, Paré ran out of oil and instead used a balm of egg yolk, rose oil and turpentine [20]. He observed that the outcomes differed substantially between the two groups: those treated with the hot oil were feverish and in ‘great pain and swelling about the edges of their wounds’, whereas those given the balm were resting comfortably [21]. Further trials of the balm convinced Paré that gunshot wounds were not poisoned and should not be cauterised [20].

The comparison of groups also helped promote a technique for the prevention of smallpox. In the 1700s smallpox was a leading cause of death, with many of those who survived suffering disfigurement and blindness [22]. The available preventive measure was to infect children with puss or scab material from smallpox victims, a process known as variolation. Despite reports that it was beneficial [23], there was widespread concern that variolation might carry a greater risk of dying than allowing people to contract the disease naturally. James Jurin evaluated this in the 1720s, by collecting data on death rates in three groups: those who were diagnosed with smallpox, those at risk of contracting smallpox and those who had been variolated [22, 23]. The results appeared convincing with death rates of 16.5% (diagnosed cases), 8.3% (at risk) and 2.0% (variolated) [23]. Preventing smallpox was a much safer practice than letting nature take its course.

Death following childbirth was a serious concern in the seventeenth to nineteenth centuries, causing epidemics ‘of unimaginable proportions’ [24]. A major cause of this mortality, puerperal fever (fever following childbirth), was investigated by Ignaz Semmelweis, a Hungarian doctor. In 1844 he compared the death rates among patients in two wards of a hospital in Vienna. He found that the death rates in a ward staffed by doctors was much higher (16%) than in the one run by midwives (2%) [25]. This, and other observations, led Semmelweis to conclude that the illness was transmitted by

doctors coming directly from a post-mortem to help deliver a baby. He initiated a preventive measure, compulsory hand washing in a chloride of lime solution, which reduced the mortality in the doctors' ward to 3% [25]. His approach was not popular, because it implied that doctors transmitted disease, and Semmelweis's contract was not renewed. He was finally vindicated some 30 years later when Pasteur identified the bacterium, *Streptococcus pyogenes*, that caused puerperal fever [25].

These treatment evaluations utilised two different types of comparisons: contemporary controls and historical controls. Contemporary controls are patients who were seen at the same time as those getting the new treatment, but who received the conventional care. Historical controls are patients who had been treated previously in the same location (e.g. hospital). Jurin's comparisons of groups at risk of smallpox, and Semmelweis's comparison of puerperal fever in two wards, used contemporary controls. In contrast the comparison of puerperal fever before and after introducing handwashing, and Paré's comparison of treatments for gunshot wounds, used historical control groups.

The problem with both types of control groups is that there could be systematic differences between the patients in the different groups. Isaac Massey, a contemporary of Jurin, made this criticism of the work on smallpox, pointing out that those who could afford variolation may have been in better health than those in the comparison groups [22]. He concluded that what was needed was groups that were similar, they 'must and ought to be as near as may be on a Par' [22].

COMPARING SIMILAR GROUPS

When groups are similar at baseline, it is more likely that any differences in subsequent outcomes will be due to the differences in the effects of the treatments. The idea of comparing like with like was proposed in the fourteenth century by the poet Francisco Petrarch, who suggested using similar groups of patients to compare the then current treatments with simply letting nature take its course [26].

One way to create similar groups is to recruit a number of patients who are all alike, then give them different treatments. The testing of potential treatments for scurvy is a widely cited example of the benefit of using similar groups. Scurvy is a debilitating and sometimes fatal disease, which afflicted sailors on long-distance sea voyages from the fifteenth to the nineteenth centuries [27, 28]. By the late 1500s, the benefits of consuming oranges and lemons were well known by Dutch sailors [27], but many English expeditions continued to suffer serious loss of life through scurvy [28]. The issue was still unresolved in 1747 when James Lind, a Royal Navy surgeon, carried out a classic study to assess the effects of six common treatments. He identified 12 sailors with scurvy who were ‘as similar as I could have them’, and tested each of the treatments on groups of 2 men (each pair to receive either: oil of vitriol, vinegar, sea water, cider, oranges and lemons, or a herbal paste) [29]. After 14 days Lind observed ‘the most sudden and visible good effects were perceived from the use of oranges and lemons’. These findings were not widely accepted, and even Lind had doubts about them [29, 30], but the method used reflects an advance in thinking about ways to test treatments. Lind is rightly celebrated for his comparison of like with like in the evaluation of treatments. (In his ‘Treatise of the Scurvy’ Lind does not make any clear recommendations for the treatment of the disease, possibly because he believed that scurvy was not due to poor diet, but was a result of faulty digestion exacerbated by wet weather [29, 30].)

Another study in the eighteenth century used similar groups to assess whether the adverse effects of variolation (to prevent smallpox) could be ameliorated by pretreatment with a compound of mercury. At that time about 1 out of 50 patients vaccinated against smallpox died following the procedure [31]. In 1767 William Watson recruited 31 children who were similar in age, gender and diet [32]. These were divided into three groups, which received either the mercury mixture, a mild senna laxative or no treatment. No clear difference was found between the groups, using an objective measure of assessment (the number of pock marks caused by the variolation). Watson concluded that variolation against smallpox was effective with or without pretreatment with mercury or a mild laxative [32].

CASTING LOTS AND TREATMENT ALLOCATION

Comparing similar groups of patients was an important step forward in the evaluation of treatments, but it leaves open the possibility that the groups may have differed on important factors that were not measured. Further, a subconscious bias in the doctor allocating patients to treatments could influence the way individuals were assigned to groups (e.g. the slightly sicker ones might be preferentially assigned to one group). An alternative approach, which prevents this bias, is to allocate individuals to treatments in a truly random way, so that the final groups will be balanced on all factors, whether measured or not.

The idea that some form of randomisation should be used to allocate patients to treatment groups was proposed in the 1640s. Joan Baptista van Helmont, a Flemish chemist, alchemist and physician, recommended this method to evaluate the effectiveness of bloodletting [33]. He suggested dividing up to 500 patients into 2 groups, then casting lots (equivalent to tossing a coin) to decide which group would be given the conventional therapy (bloodletting) and which would receive van Helmont's own treatment. A notable feature of the trial design is that the outcome would be decided by the number of funerals that occurred in the two groups. The experiment was not carried out. (The proposed use of an objective outcome measure such as this is unusual for its time.)

One method of randomised allocation was used in 1848 by Thomas Graham Balfour to investigate whether homeopathic belladonna could prevent scarlet fever. Balfour identified 151 boys who had not had the disease, and 'divided them into two sections, taking them alternately from the list, to prevent the imputation of selection' [34]. Balfour recognised that if he had to decide which boys were allocated to each group, his choices might be biased. (Alternate selection from a list is essentially a method of randomisation, as the factors which are related to dying from scarlet fever, will be randomly scattered throughout the list.) The study showed that exactly two children in each group developed scarlet fever, leading him to conclude that 'the numbers are too small to justify deductions as to the prophylactic power of belladonna' [34], a commendably careful interpretation of the findings.

Instead of alternate selection from a list, patients could be allocated to treatments by the date of their admission to hospital. This method was used by the Danish physician Johannes Fibiger in 1896–1897 [35] to evaluate the effectiveness of a serum treatment for diphtheria. Thus, patients admitted to hospital on one day received serum and those on the next day were untreated. The outcome was persuasive: only 8 of 239 patients in the serum group died, compared to 30 of the 245 controls.

The use of alternate allocation began to gain popularity in the first few decades of the twentieth century because it prevented bias in the assignment of patients to treatments. These research studies were conducted in both the United States and the UK, with patients being randomised by the order of their attendance at a healthcare facility [36–39]. These trials signalled the growing recognition of the importance of achieving comparable groups.

RANDOM NUMBERS FOR TREATMENT ALLOCATION

A landmark series of three trials, conducted under the auspices of the UK Medical Research Council, used random numbers to allocate patients to treatments. This methodological advance was proposed by the medical statistician Professor (later Sir) Austin Bradford Hill [40]. It was first used in a large field trial that assessed the effectiveness of a vaccine for whooping cough [41]. Although this study began in 1944, it was not published until 1951. The second trial, of streptomycin for pulmonary tuberculosis, became the most highly acclaimed study in the history of treatment evaluation. It began in 1946, but was the first to be published, in 1948 [42]. The third trial involved a large-scale field trial of an antihistaminic drug (thonzylamine) for the prevention of the common cold [43].

As well as being published first, the streptomycin trial provided a major advance in the treatment of a feared disease, tuberculosis: it reduced the fatality rate at six months from 27% to 7% and also reduced the severity of disease among survivors. An editorial that accompanied the paper identified the advantage of individual randomisation over alternate allocation: it prevented a patient being

included or rejected, based on whether the next treatment was to be antibiotic or control [44]. For example, if the doctor thought that the drug would not be effective in seriously ill patients, they might not be included in the study if they were scheduled to receive the active treatment. This would only need to happen a few times to bias the results of the study.

In addition to the use of random numbers to allocate patients to treatments, these three trials stand out for two other reasons. Patients were recruited from multiple centres to provide sufficient participants to be able to draw firm conclusions. The researchers also made considerable efforts to ensure that the participants, and the clinicians measuring the outcomes, were unaware of which treatment the patients received. This prevented bias in the reporting of symptoms by participants, and by those recording the outcomes: in modern terminology, it was double blind.

The landmark streptomycin trial in tuberculosis was followed by another study on pulmonary tuberculosis, published two years later [45]. It compared three treatments: streptomycin, another drug, para-amino-salicylic acid (PAS), and a combination of these two drugs. The same methodology was used as in the first streptomycin trial. The combination therapy had the best outcome, with streptomycin coming second. More importantly the combined treatment led to a much lower frequency of bacterial resistance to streptomycin. This study has been credited with leading to the maxim ‘never treat active tuberculosis with a single agent’, which is now the standard for managing this disease [46]. The clinical benefits apart, this set of four rigorous studies supported by the Medical Research Council inaugurated the era of high quality clinical trials.

THE NEED FOR BLINDING

A major concern of several trials in the middle of the twentieth century was to ensure that the patients and clinical observers were unaware of how treatments were allocated [39, 41, 43]. This would prevent knowledge of who received which treatment from influencing the outcome of the study. The idea was not new; it featured in studies to

evaluate a treatment called animal magnetism. This treatment was championed in the late 1700s by Franz Anton Mesmer, who believed he could impart magnetic energy and thereby cure a wide range of illnesses [47]. Mesmer achieved great fame, and a lucrative medical practice in Paris. This popularity was of such concern to other doctors, and to the government, that they persuaded Louis XVI to establish a Royal Commission to evaluate the claims of cures and dramatic effects [47, 48]. The Commission conducted a series of studies in which participants either thought they were being magnetised (when they were not), or thought they were not subjected to magnetism (when they were). The findings were convincing. Participants only reported benefits when they (falsely) believed they were being treated: ‘the imagination is the real cause of the effects attributed to magnetism’ [47]. Following publication of the Commission’s report, Mesmer was ridiculed, and animal magnetism was abandoned in France.

A more recent example of the importance of blinding is the evaluation of a surgical technique, internal-mammary artery-ligation, for the relief of angina symptoms. Several reports in the 1950s had claimed that the operation provided considerable relief of symptoms [49, 50]. This prompted two groups of researchers to carry out controlled trials to evaluate the effectiveness of the surgery. Patients were randomly allocated to have artery ligation, or to a control group which received a sham operation involving only a skin incision. The patients, and the cardiologists who evaluated the outcomes, were blind to treatment group. The ligation operation provided no benefit, as most patients in both the treatment group and in the control group reported significant improvement in symptoms [49, 50]. The authors concluded that these claims were most likely a psychological response to undergoing surgery.

The response to a sham treatment is known as a placebo effect. Understanding of the psychological and physiological factors underlying the placebo response has advanced greatly in recent years [51, 52]. A consistent finding is that patients who have high expectations of their treatment usually experience improvements in symptoms. If patients were aware of their treatment allocation, only those in the active group would have the high expectations. Concealment of treatment allocation could prevent this bias from creating difference between the groups.

CONCLUSION

Obtaining evidence on treatment effectiveness is a challenging business. As Passamani remarked in 1991, 'The history of medicine is richly endowed with therapies that were widely used and then shown to be ineffective or frankly toxic' [53]. A similar view was expressed by the celebrated American physician, Oliver Wendell Holmes in 1860, 'if the whole materia medica, as now used, could be sunk to the bottom of the sea, it would be all the better for mankind – and all the worse for the fishes' [54]. These may seem somewhat jaundiced views, but they reflect the large proportion of ineffective and possibly harmful treatments that were once used. Even in the early years of the twentieth century many ineffective treatments were widely used [55], and some treatments of little value continue to be used today [56]. Concern about this has led to a recent international campaign, 'Choosing Wisely', to reduce the use of ineffective or harmful treatments [57].

This chapter has presented examples of different approaches used to identify potentially effective treatments. Reliance on theories of disease processes is often unreliable and can result in harmful treatments being used. Careful observation of treatment outcomes in a series of patients can, if the benefits are immediate and substantial, identify effective treatments. Comparisons of groups of patients given different treatments are often more insightful, but are vulnerable to the criticism that the groups might not be similar at baseline. As the eminent French physician P C A Louis pointed out in 1834, 'it is necessary to account for differences of age, sex, temperament, physical condition, natural history of the disease' [58]. The use of groups constructed to be similar on some factors at baseline is a definite improvement, but leaves open the question that they differ on other (unmeasured) factors. Allocation of individual patients to treatments using random numbers overcomes two problems: clinician bias in assigning patients to groups, and differences in unmeasured factors.

The sequence of methods presented in this chapter could be taken to imply that there was a steady progression to increased robustness of study design. However, as the dates for the individual

studies show, there is little evidence for continuous improvement in methods: rather there was substantial overlap in the use of these methods. The major advances in trial methodology occurred in studies conducted in the middle of the twentieth century. They used three techniques that are now hallmarks of high quality trials: randomisation, blinding of the investigators and patients to the randomisation process, and objective outcome measures.

In summary, this chapter has reviewed the development of methods to evaluate treatments up to the middle of the twentieth century. It has highlighted pitfalls of many of the earlier methods and concluded with an outline of the advantages of the double blind randomised controlled trial. This method is now used around the world to identify the benefits of treatments. Medicine now has the tools to ensure that only effective treatments are used. The next chapter explores whether the benefits of the randomised controlled trial have been realised.

REFERENCES

1. Porter, R. (1999). *The Greatest Benefit to Mankind*. London: Harpers Collins.
2. Cantu, J.Q. (1965). Benjamin Franklin's medical imprints. *Bull. Med. Libr. Assoc.* 53: 71–79.
3. Bhatt, A. (2010). Evolution of clinical research: a history before and beyond Lames Lind. *Perspect. Clin. Res.* 1: 6–10.
4. Ackerknecht, E.H. (1982). *A Short History of Medicine*. Baltimore: Johns Hopkins University Press.
5. DePalma, R.G., Hayes, V.W., and Zacharski, L.R. (2007). Bloodletting: past and present. *J. Am. Coll. Surg.* 205: 132–144.
6. Risse, G.B. (1979). Renaissance of bloodletting – chapter in modern therapeutics. *J. Hist. Med. Allied Sci.* 34: 3–22.
7. Stokes, W. (1865). The address in medicine. *BMJ* ii: 133–142.
8. Biggio, J.R. Jr. (2013). Bed rest in pregnancy: time to put the issue to rest. *Obstet. Gynecol.* 121: 1158–1160.
9. Sprague, A.E. (2004). The evolution of bed rest as a clinical intervention. *J. Obstet. Gynecol. Neonatal. Nurs.* 33: 542–549.

10. Bigelow, C. and Stone, J. (2011). Bed rest in pregnancy. *Mt Sinai J. Med.* 78: 291–302.
11. McCall, C.A., Grimes, D.A., and Lyerly, A.D. (2013). ‘Therapeutic’ bed rest in pregnancy: unethical and unsupported by data. *Obstet. Gynecol.* 121: 1305–1308.
12. Brower, R.G. (2009). Consequences of bed rest. *Crit. Care Med.* 37: S422–S428.
13. Jack, D.B. (1997). One hundred years of aspirin. *Lancet* 350: 437–439.
14. Montinari, M.R., Minelli, S., and De Caterina, R. (2019). The first 3500years of aspirin history from its roots – a concise summary. *Vasc. Pharmacol.* 113: 1–8.
15. Vane, J.R., Flower, R.J., and Botting, R.M. (1990). History of aspirin and its mechanism of action. *Stroke* 21: IV12–IV23.
16. Rosenfeld, L. (2002). Insulin: discovery and controversy. *Clin. Chem.* 48: 2270–2288.
17. Banting, F.G. (1926). An address on diabetes and insulin: being the Nobel lecture delivered at Stockholm on September 15th, 1925. *CMAJ* 16: 221–232.
18. Banting, F.G., Best, C.H., Collip, J.B. et al. (1922). Pancreatic extracts in the treatment of diabetes mellitus. *CMAJ* 12: 141–146.
19. Karamitsos, D.T. (2011). The story of insulin discovery. *Diabetes Res. Clin. Pract.* 93 (Suppl 1): S2–S8.
20. Wangenstein, O.H., Wangenstein, S.D., and Klinger, C.F. (1972). Wound management of Ambroise pare and Dominique Larrey, great French military surgeons of the 16th and 19th centuries. *Bull. Hist. Med.* 46: 207–234.
21. Drucker, C.B. (2008). Ambroise pare and the birth of the gentle art of surgery. *Yale J. Biol. Med.* 81: 199–202.
22. Huth, E. (2006). Quantitative evidence for judgments on the efficacy of inoculation for the prevention of smallpox: England and New England in the 1700s. *J. R. Soc. Med.* 99: 262–266.
23. Bird, A. (2019). James Jurin and the avoidance of bias in collecting and assessing evidence on the effects of variolation. *J. R. Soc. Med.* 112: 119–123.

24. Charles, D. and Larsen, B. (1986). Streptococcal puerperal sepsis and obstetric infections: a historical perspective. *Rev. Infect. Dis.* 8: 411–422.
25. Adriaanse, A.H., Pel, M., and Bleker, O.P. (2000). Semmelweis: the combat against puerperal fever. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 90: 153–158.
26. Chalmers, I., Dukan, E., Podolsky, S. et al. (2012). The advent of fair treatment allocation schedules in clinical trials during the 19th and early 20th centuries. *J. R. Soc. Med.* 105: 221–227.
27. Burnby, J. and Bierman, A. (1996). The incidence of scurvy at sea and its treatment. *Rev. Hist. Pharm.* 44: 339–346.
28. Baron, J.H. (2009). Sailors' scurvy before and after James Lind – a reassessment. *Nutr. Rev.* 67: 315–332.
29. Milne, I. (2012). Who was James Lind, and what exactly did he achieve. *J. R. Soc. Med.* 105: 503–508.
30. Bartholomew, M. (2002). James Lind's treatise of the scurvy (1753). *Postgrad. Med. J.* 78: 695–696.
31. Boylston, A.W. (2002). Clinical investigation of smallpox in 1767. *N. Engl. J. Med.* 346: 1326–1328.
32. Boylston, A. (2014). William Watson's use of controlled clinical experiments in 1767. *J. R. Soc. Med.* 107: 246–248.
33. Donaldson, I.M. (2016). Van Helmont's proposal for a randomised comparison of treating fevers with or without bloodletting and purging. *J. R. Coll. Physicians Edinb.* 46: 206–213.
34. Chalmers, I. and Toth, B. (2009). Nineteenth-century controlled trials to test whether belladonna prevents scarlet fever. *J. R. Soc. Med.* 102: 549–550.
35. Hrobjartsson, A., Gotzsche, P.C., and Gluud, C. (1998). The controlled clinical trial turns 100 years: Fibiger's trial of serum treatment of diphtheria. *BMJ* 317: 1243–1245.
36. Podolsky, S.H. (2009). Jesse Bullowa, specific treatment for pneumonia, and the development of the controlled clinical trial. *J. R. Soc. Med.* 102: 203–207.
37. MRC (1934). The serum treatment of lobar pneumonia: a report of the therapeutic trials Committee of the Medical Research Council. *BMJ* 1: 241–245.

38. Lorriman, G. and Martin, W.J. (1950). Trial of antistin in the common cold. *BMJ* 2: 430–431.
39. MRC (2004). Clinical trial of patulin in the common cold. 1944. *Int. J. Epidemiol.* 33: 243–246.
40. Hill, A.B. (1990). Suspended judgment. Memories of the British streptomycin trial in tuberculosis. The first randomized clinical trial. *Control. Clin. Trials* 11: 77–79.
41. MRC (1951). Prevention of whooping-cough by vaccination; a Medical Research Council investigation. *BMJ* 1: 1463–1471.
42. MRC (1948). Streptomycin treatment of pulmonary tuberculosis. *BMJ* 2: 769–782.
43. MRC (1950). Clinical trials of antihistaminic drugs in the prevention and treatment of the common cold; report by a special committee of the Medical Research Council. *BMJ* 2: 425–429.
44. Editorial (1948). The controlled therapeutic trial. *BMJ* 2: 791–792.
45. MRC (1950). Treatment of pulmonary tuberculosis with streptomycin and Para-aminosalicylic acid; a Medical Research Council investigation. *BMJ* 2: 1073–1085.
46. Murray, J.F., Schraufnagel, D.E., and Hopewell, P.C. (2015). Treatment of tuberculosis. A historical perspective. *Ann. Am. Thorac. Soc.* 12: 1749–1759.
47. Lanska, D.J. and Lanska, J.T. (2007). Franz Anton Mesmer and the rise and fall of animal magnetism: dramatic cures, controversy, and ultimately a triumph for the scientific method. In: *Brain, Mind and Medicine: Essays in Eighteenth-Century Neuroscience* (eds. H. Whitaker, C.U.M. Smith and S. Finger), 301–320. New York: Springer <https://doi.org/10.1007/978-0-387-70967-3>.
48. Donaldson, I.M. (2005). Mesmer’s 1780 proposal for a controlled trial to test his method of treatment using “animal magnetism”. *J. R. Soc. Med.* 98: 572–575.
49. Dimond, E.G., Kittle, C.F., and Crockett, J.E. (1960). Comparison of internal mammary artery ligation and sham operation for angina pectoris. *Am. J. Cardiol.* 5: 483–486.
50. Cobb, L.A., Thomas, G.I., Dillard, D.H. et al. (1959). An evaluation of internal-mammary-artery ligation by a double-blind technic. *N. Engl. J. Med.* 260: 1115–1118.

-
51. Enck, P., Bingel, U., Schedlowski, M. et al. (2013). The placebo response in medicine: minimize, maximize or personalize? *Nat. Rev. Drug Discov.* 12: 191–204.
 52. Horing, B., Weimer, K., Muth, E.R. et al. (2014). Prediction of placebo responses: a systematic review of the literature. *Front. Psychol.* <https://doi.org/10.3389/fpsyg.2014.01079>.
 53. Passamani, E. (1991). Clinical trials – are they ethical? *N. Engl. J. Med.* 324: 1589–1592.
 54. Schimmel, E.M. (1963). The physician as pathogen. *J. Chronic Dis.* 16: 1–4.
 55. Beeson, P.B. (1980). Changes in medical therapy during the past half century. *Medicine* 59: 79–99.
 56. Mafi, J.N. and Parchman, M. (2018). Low-value care: an intractable global problem with no quick fix. *BMJ Qual. Saf.* 27: 333–336.
 57. Levinson, W., Kallewaard, M., Bhatia, R.S. et al. (2015). ‘Choosing Wisely’: a growing international campaign. *BMJ Qual. Saf.* 24: 167–174.
 58. Best, M. and Neuhauser, D. (2005). Pierre Charles Alexandre Louis: master of the spirit of mathematical clinical science. *Qual. Saf. Health Care* 14: 462–464.