Why the 1948 MRC trial of streptomycin used treatment allocation based on random numbers

Iain Chalmers
James Lind Initiative, Summertown Pavilion, Middle Way, Oxford OX2 7LG, UK
Email: ichalmers@jameslindlibrary.org

There is a view among some medical historians that the emergence of the randomized clinical trial originated from statistical thinking, and that the modern era of controlled trials was essentially ushered in with the iconic randomized trial of streptomycin for pulmonary tuberculosis reported by the British Medical Research Council (MRC) in 1948. For example:

The professional emergence of statistics as a codified body of knowledge and the concomitant rise of individuals trained in its methods provided the necessary conditions for the Laplacian vision of the probabilistically based clinical trial to come into being.\(^1\)

The randomized clinical trial is ‘an extension of the statistician RA Fisher’s ideas about experimental design’ (p. 132). ‘The statisticians’ randomized controlled trial came to represent the symbol and substance of the statistical method in medicine’.\(^2\)

The conceptualization of clinical trials as ‘a seminal statistical idea’ which ‘can be traced back to RA Fisher’s work’ has not been demonstrated by these writers or by others. The early history of clinical trials has little to do with statistical theory and much more to do with the more fundamental and less technical concept of a fair – that is, unbiased – test.\(^5\–\(^11\)

The need to ‘compare like with like’ in fair tests of treatments has been recognized by some people for a long time, and not only by physicians. In a letter to Boccacio written in 1364, Petrarch wrote:

I solemnly affirm and believe, if a hundred or a thousand 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 half followed the prescriptions of the doctors of the variety of those practising 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.\(^12\)

When quantitative methods began to be used at the beginning of the 18th century to assess the effects of variolation authors of the comparisons were sometimes reminded of the need to ensure that like was being compared with like. Thus Massey, challenging the interpretation of comparisons of mortality following variolation and after natural smallpox, wrote:

…to form a just Comparison, and calculate right in this Case, the Circumstances of the Patients, must and ought to be as near as may be on a Par.\(^13\)

Several reports of prospective experiments were published during the 18th century. In the most celebrated of these James Lind notes that, apart from the treatments, the 12 patients he studied were otherwise similar: ‘They all in general had putrid gums, the spots and lassitude, with
weakness of their knees. They lay together in one place, being a proper apartment for the sick in the fore-hold; and had one diet common to all.\textsuperscript{14} Lind does not tell us how he allocated his 12 patients to each of the six treatments he compared, but had he cast lots or used alternation or rotation it would not have been inconsistent with the use of these devices to make fair decisions in other contexts.\textsuperscript{15}

At the beginning of the 19th century, Alexander Hamilton reported having used alternation to generate parallel comparison groups in a clinical trial of bloodletting done by him and two surgeon colleagues.\textsuperscript{16} He described how sick soldiers had been ‘admitted, \textit{alternately} [my emphasis], in such a manner that each of us had one third of the whole’, and that ‘the sick were indiscriminately received’, and ‘attended as nearly as possible with the same care and accommodated with the same comforts’.\textsuperscript{16} Although his report leaves several uncertainties,\textsuperscript{17} it seems reasonable to speculate that he described the use of alternation to show that an effort had been made to generate comparable treatment groups.

By the middle of the 19th century, the rationale for alternation was sometimes being made explicit. In 1854, Thomas Graham Balfour described his assessment of whether belladonna could prevent scarlet fever. He divided 151 boys into two comparison groups, ‘taking them alternately from the list, \textit{to avoid the imputation of selection} [my emphasis].’\textsuperscript{18} It is clear from these words that Balfour used alternation to control selection bias. This is not a statistical concept, and although Balfour was a distinguished statistician as well as a doctor, he cannot be regarded as a theoretical statistician in the ‘Pearsonian/Fisherian’ sense.\textsuperscript{19}

There are further isolated examples of alternation being used to generate treatment comparison groups during the last half of the 19th century, but they became increasingly common during the first half of the 20th century. Indeed, alternation as a feature of research design became referred to formally in English not only simply as ‘alternation’,\textsuperscript{20} but also as ‘the alternate method’, ‘rational alternation’,\textsuperscript{21} and ‘the alternate case method’.\textsuperscript{21,22} In French it was referred to as ‘la méthode alternante’,\textsuperscript{23,24} and in German as ‘Simultanmethode’.\textsuperscript{25} It is worth noting that designation of this methodological principle occurred before the theoretical statistical qualities of random allocation had been promoted in Ronald Fisher’s \textit{The Design of Experiments}.\textsuperscript{26} Indeed, even though the word ‘random’ sometimes appeared in reports of controlled trials before the late 1940s, it was often actually alternation that was being used for allocation.\textsuperscript{27}

Unsurprisingly therefore, the use of alternation was reflected in articles and a book published by the \textit{Lancet} in 1937, written by the father of medical statistics in Britain, Austin Bradford Hill:

\begin{quote}
By the allocation of the patients to the two groups we want to ensure that these two groups are alike except in treatment… this might be done, with reasonably large numbers, by a random division of the patients; the first being given treatment A, the second being orthodoxy treated and serving as a control, the third being given treatment A, the fourth serving as a control, and so on, no departure from this rule being allowed [my emphasis].\textsuperscript{28}
\end{quote}

Of the two essential components of unbiased allocation – genesis of an unbiased sequence, and unbiased implementation of the sequence – the former remains a trivially easy task, while the latter will continue to pose challenges.\textsuperscript{11} Hill was aware of this. In an internal report for the MRC dated 22 December 1933, Hill expressed concern about the allocation of patients to comparison groups in a MRC study of serum treatment for pneumonia in which alternation should have been used.\textsuperscript{29} Imbalance in the sizes of the comparison groups made clear that alternation had not been strictly observed, prompting Hill to stress the theoretical statistical qualities of random allocation had been promoted in Ronald Fisher’s \textit{The Design of Experiments}.\textsuperscript{26} Indeed, even though the word ‘random’ sometimes appeared in reports of controlled trials before the late 1940s, it was often actually alternation that was being used for allocation.\textsuperscript{27}

This principle was reflected in the first properly controlled multicentre trial conducted under the aegis of the British MRC. This was designed by Philip D’Arcy Hart to assess the effect of patulin on common cold symptoms.\textsuperscript{30–32} When I interviewed him 60 years later, he told me:

\begin{quote}
Everyone had thought we would use alternation, and we thought we were very clever in setting up a scheme with two patulin groups and two placebo groups using letters to designate each of the four
\end{quote}
groups, then using rotation to allocate people to the different groups. We thought we were doing something completely new. We wanted to muddle people up. In fact we succeeded in muddling ourselves up. We didn’t always remember what the letters stood for. None of us was a statistician, but we felt that the patulin trial was the first decently controlled trial the MRC had done. (IC interview with Philip D’Arcy Hart, 2 May 2003)

D’Arcy Hart was one of the team – with Marc Daniels and Austin Bradford Hill – that designed the MRC streptomycin trial. The report of the study is a model of clarity. A crucially important element is the statement that ‘the details of the (allocation) series were unknown to any of the investigators or to the coordinator and were contained in a set of sealed envelopes, each bearing on the outside only the name of the hospital and a number’.33 The reason that the MRC streptomycin trial deserves its place in the history of clinical trials is this and other exceptionally clear statements assuring readers that adequate precautions had been taken to minimize the possibility of allocation bias, and thus assure readers that ‘like would be compared with like’.34,35

In spite of a few examples of random allocation during the 1920s and 1930s, alternation remained the principal method for unbiased prospective allocation to treatment comparison groups36,37 until well after the end of World War II, even in studies done by investigators such as Richard Doll, who were very familiar with Fisher’s writings.38 The ‘clinical’ and ‘statistical’ reasons for random allocation came together only during the second half of the 20th century. But even today, as has been noted by the distinguished statistician David Cox, the primary reason for using random allocation is not statistical, but to help prevent foreknowledge of treatment assignments, and thus the conscious or unconscious temptation to allow biased allocation to occur.39

References
2 Marks HM. The Progress of Experiment. Cambridge: Cambridge University Press, 2000
6 Chalmers I. Why transition from alternation to randomisation in clinical trials was made. BMJ 1999;319:1372
7 Chalmers I. Comparing like with like: some historical milestones in the evolution of methods to create unbiased comparison groups in therapeutic experiments. Int J Epidemiol 2001;30:1156–64
10 Edwards MV. Control and the therapeutic trial. Amsterdam: Rodopi, 2006
11 Chalmers I. Explaining the unbiased creation of treatment comparison groups. Lancet 2009;374:1670–1
12 Petrarca F. Rerum Senilium Libri. Liber XIV, Epistola I. Letter to Boccaccio (V3). 1364
13 Massey I. A short and plain account of inoculation. With some remarks on the main argument made use of to recommend that practice, by Mr. Maitland and others. To which is added, a letter to the learned James Jurin, M.D.R.S. Secr. Col. Reg. Med. Lond. Soc. In answer to his letter to the learned Dr. Cotesworth, and his comparison between the mortality of natural and inoculated small pox. The second edition. London: W Meadows, 1723
14 Lind J. A treatise of the scarify. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronologial view of what has been published on the subject. Edinburgh: A Kincaid and A Donaldson, 1753
16 Hamilton AL. Dissertatio medica inauguralis de synocho castrens (Inaugural medical dissertation on camp fever). Edinburgh: J Ballantyne, 1816
18 Balfour TG. Quoted in West C. Lectures on the diseases of infancy and childhood. London: Longman, Brown, Green and Longmans, 1854
20 Bullowa JGM. The use of antipneumococcic refined serum in lobar pneumonia: data necessary for a comparison between cases treated with serum and cases not so treated, and the importance of a significant control series of cases. JAMA 1928;90:1354–8

Downloaded from jrsm.sagepub.com by guest on November 8, 2016
21 Choksy KBNH. On recent progress in serum-therapy of plague. BMJ 1908;1:1282–4
22 Cecil RL, Plummer N. Pneumococcus Type I pneumonia - a study of eleven hundred and sixty-one cases, with especial reference to specific therapy. JAMA 1930;95:1547–53
26 Fisher RA. The design of experiments. Edinburgh: Oliver and Boyd, 1935
31 Clarke M. The 1944 patulin trial of the British Medical Research Council: an example of how concerted common purpose can get reliable answers to important questions very quickly. The James Lind Library 2004. See www.jameslindlibrary.org