

Statistical Theory Was Not the Reason That Randomization Was Used in the British Medical Research Council's Clinical Trial of Streptomycin for Pulmonary Tuberculosis

IAIN CHALMERS

**"RANDOMIZATION WAS INTRODUCED
TO CONTROL SELECTION BIASES,
NOT FOR ANY ESOTERIC STATISTICAL REASON."¹**

An understanding of the history of controlled trials is of importance today because their history is still evolving. Although it is possible to identify treatments with dramatic effects – good and bad – without carefully controlled studies, inferences about more typical treatment effects are usually insecure unless based on studies with concurrent comparison groups, assembled in ways that reduce the likelihood that biases or the play of chance will mislead people.

The principal defining characteristics of controlled clinical trials today are the measures taken to reduce biases and the play of chance. Although several historians have been interested in "the taming of chance" – Ian Hacking's apt expression² – very few have focused on "beating biases."³ Recent exceptions include Kaptchuk's history of the evolution of measures to reduce observer biases in clinical trials⁴ and Harry Marks's commentary on some aspects of control of biases in clinical trials after 1950.⁵

The apparent lack of interest in bias by historians of clinical trials is particularly surprising, given the importance that they ascribe to the 1948 report of Britain's Medical Research Council (MRC) of a randomized clinical trial of streptomycin in pulmonary tuberculosis.⁶ The

report is notable for its exceptionally clear description of measures taken to control bias, and it rapidly became a historical landmark. At the annual meeting of the American Association of the History of Medicine in 1954, Donald Mainland, after describing the many problems that faced him as a medical statistician, declared: "all has not been darkness ... In the clinical field there appeared in 1948 a beacon or lighthouse beam – the report of the British Medical Research Council's co-operative trial of streptomycin in pulmonary tuberculosis."⁷

The report noted that allocation of patients to the comparison groups had been accomplished by reference to "a statistical series based on random sampling numbers drawn up for each sex at each centre by Professor Bradford Hill."⁸ These words reflect the statistical language developed by theorists such as Karl Pearson and Ronald Fisher during the previous half-century. Against a background of interest in the history of probability and inferential statistics, some historians believe that the evolution of statistical theory played a key role in the evolution of the randomized clinical trial. Rosser Matthews,⁹ for example, suggests: "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." Harry Marks¹⁰ judges the randomized clinical trial to have been "an extension of the statistician R.A. Fisher's ideas about experimental design" and that "the statisticians' randomized controlled trial came to represent the symbol and substance of the statistical method in medicine."¹¹ Jean-Paul Gaudillière¹² observes: "The history of randomized clinical trials may be traced back to the biometricians' work and it seems to be a good example of 'applied statistics'. On the one hand there was a direct lineage from Pearson to Bradford Hill via Fisher and Major Greenwood ... On the other hand, it is not too difficult to argue for conceptual legacy, since the basic concepts grounding the choice of randomisation can be traced back to R.A. Fisher's work." Most recently, Eileen Magnello has stated that Karl Pearson's 1904 proposal for a clinical trial using alternation to generate the comparison groups constituted a "seminal statistical idea."¹³

I have been unable to find any evidence to support these interpretations of the origins of the MRC's randomized trial of streptomycin, and I propose an alternative history. This 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, which is a separate concern in the history of ideas about study design.

To defend this proposition, I begin by describing the two separate steps needed to achieve unbiased allocation to treatment groups in

clinical trials'— generating unbiased allocation schedules intended to ensure that like will be compared with like; and preventing foreknowledge of allocations among those involved in recruiting patients to clinical trials. I show that formal random allocation coexisted with alternate allocation in medicine throughout most of the twentieth century; that the word 'random' has often been used loosely, without any necessary link to the significance of random allocation in statistical theory; and that random allocation was adopted for the MRC trial of streptomycin for pulmonary tuberculosis to prevent foreknowledge of allocations among those involved in recruiting patients. I draw extensively on the writings, official and unofficial, attributed and unattributed, of the main protagonist, Austin Bradford Hill, and compare the relevant passages in successive editions of his textbook. I end by noting that the 'clinical' and 'statistical' reasons for random allocation came together only during the second half of the twentieth century.

UNBIASED ALLOCATION TO COMPARISON GROUPS:
TWO SEPARATE STEPS, BOTH ESSENTIAL

Assembling comparison groups in clinical trials such that any differences in measured *and unmeasured* variables of prognostic importance are due solely to chance involves two quite separate steps, both of which are essential to ensure comparison of like with like.¹⁴

Generating Allocation Schedules

The first of the two steps involves using an unbiased method to decide which of the comparison groups each patient will join. One may generate allocation schedules using alternation or rotation, by tossing coins or drawing lots, or by reference to tables of random sampling numbers (the method used in the 1948 streptomycin trial) or to computer-generated lists of (pseudo-)random numbers.

Contrary to widespread belief, allocating by strict alternation does not control bias less effectively than use of random numbers.¹⁵ Clearly, if some factor of possible prognostic importance confounds alternation, it will not control (allocation) bias. For example, if factors other than chance had influenced the day of the week on which Semmelweis's maternity hospital in Vienna admitted women, or the days on which Fibiger's hospital in Copenhagen admitted patients with diphtheria, then their use of "day of hospital admission" to construct comparison groups might have been biased.¹⁶

If alternation is not so correlated with potential confounders, formal random assignment (based on coin tosses, for example) controls

bias no better than strict alternation in a consecutive series of people. This shared feature of the two approaches surfaces in the frequent use of the word “random” with reference to alternate allocation to comparison groups. For example, in his *Principles of Medical Statistics*, Bradford Hill noted that alternation results in “a random division of the patients among the comparison groups in a trial,” as long as “no departure from this rule is allowed.”¹⁷ Indeed, strict alternation actually generates comparison groups that are more alike than groups formed using simple randomization.¹⁸

In the decades before and after the MRC streptomycin trial, Hill was one of many writers¹⁹ to use “random” in a sense that is less specific than the concept proposed by statistical theorists, although Fisher and other theoretical statisticians (for example, ‘Student’) debated the relative merits of alternation and randomization.²⁰

Preventing Foreknowledge of Allocations

Whether one uses alternation, randomization, or some other unbiased method to generate an unbiased allocation schedule, strict observance of the allocations generated is crucial. For this reason, it is essential to prevent foreknowledge of the allocations among clinicians, patients, and others involved in recruiting participants to trials and so prevent subversion of the allocation schedule (cheating!). As empirical research has amply demonstrated,²¹ failure to conceal allocation schedules and adherence to them will introduce bias.²²

Generating allocation schedules based on random numbers cannot – simply through the mystique of randomization – guarantee the avoidance of bias in assembling comparison groups in clinical trials. Quite apart from the “open invitation” to introduce bias that would result from pinning a random-allocation schedule on a noticeboard in a clinic where patients were being assessed for possible eligibility,²³ some “concealed” random allocations can be guessed, especially if organized within small blocks of unvarying size. Bradford Hill recognized this problem and drew attention to it in his discussion of a paper by Peter Armitage in 1959.²⁴ Of the two essential components of unbiased allocation – genesis of an unbiased sequence, and unbiased implementation of the sequence – the former remains trivially easy, while the latter will continue to pose challenges.

In the streptomycin trial “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.”²⁵ The MRC streptomycin trial deserves its place in the history of clinical trials because of this

and other exceptionally clear statements about precautions to minimize allocation bias.

THE EVOLUTION OF ALLOCATION SCHEDULES IN CLINICAL TRIALS

"Comparing like with like" in therapeutic experiments receives insufficient appreciation even today; but for at least two centuries some people have recognised its importance. In James Lind's 1753 account of his clinical trial of treatments for scurvy, for example, he notes that, apart from the treatments, the 12 patients whom 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."²⁶ Lind does not tell us how he allocated his twelve patients to each of the six treatments that 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.²⁷ Other controlled comparisons reported in the eighteenth century involved giving different treatments to the same patient for alternating periods. Thus Caleb Parry reported on a series of patients, to each of whom he had given either imported Turkish rhubarb or native rhubarbs, one after the other, to assess their relative merits as aperients.²⁸

In 1816, Alexander Hamilton reported having used alternation to generate parallel comparison groups in a clinical trial of bloodletting in 366 soldiers during the Peninsular War.²⁹ Hamilton described how sick soldiers had been "admitted, alternately" under the care of surgeons who either used or withheld venesection, but whose patients were otherwise "attended as nearly as possible with the same care and accommodated with the same comforts"³⁰. His report leaves several uncertainties,³¹ but it seems reasonable to speculate that he described the use of alternation to show that he had tried to generate comparable treatment groups.

By the mid-nineteenth century, the rationale for alternation was sometimes 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, to avoid the imputation of selection" (emphasis added).³² Balfour clearly used alternation to control bias. Although he was a distinguished statistician as well as a doctor, however, he was not a theoretical statistician in the 'Pearsonian/Fisherian' sense.³³

There are further isolated examples of alternation up to the mid-1920s,³⁴ in addition to Pearson's unfulfilled proposal for an alternate

allocation trial.³⁵ From the late 1920s on, however, there were more and more reports of alternation in clinical trials, with well over 30 published by 1948.³⁶ Among these, at least four took place under the aegis of the MRC's Therapeutic Trials Committee.³⁷ The first multi-centre, placebo-controlled trial under the MRC's aegis was organized by its Patulin Clinical Trials Committee in 1944.³⁸ The committee was chaired by Harold Himsworth (later head of the MRC) and included three fellows of the Royal Society, among them the medical statistician Major Greenwood. A system of "strict rotation" allocated volunteers suffering from the common cold to comparison groups. Committee Secretary Philip D'Arcy Hart observed that the method ensured "an effectively random allocation of the subjects to patulin and placebo."³⁹ Despite the reported use of random sampling numbers in the MRC's streptomycin trial, alternation continued, even in MRC trials,⁴⁰ and remains in use today.⁴¹

As already noted above, the term "random" has been and continues to be used quite loosely, without any necessary conceptual reference to its technical meaning in statistical theory. It seems improbable in the extreme, for example, that the seventeenth-century Flemish physician Van Helmont had statistical theory in mind when he proposed casting lots to decide which patients should be treated by orthodox medical practitioners, using bloodletting and purging, and which patients he should treat without these unpleasant interventions.⁴² Van Helmont's proposal to cast lots to decide who would have which patients almost certainly reflected his belief that this was a way of ensuring a fair therapeutic contest, in the same way that lots had been cast over centuries to make other fair decisions.⁴³

Investigators have reported drawing lots to generate comparison groups in clinical research since at least August 1927. In that month, recruitment began in an MRC-controlled trial of the effects of ultraviolet radiation on the health and development of children. The youngsters in each of seven classes were divided into three groups by "drawing lots (*method not specified*) ... so that the three large groups should be composed of children of the same ages whose school life was influenced by similar conditions."⁴⁴ Two years later, James Doull and his colleagues studied the effects of ultraviolet radiation on the health of staff members and students at Johns Hopkins.⁴⁵ Professor Lowell Reed, a mathematician colleague of the American medical statistician Raymond Pearl, allocated volunteers to the comparison groups using different-coloured dice. These were "thoroughly mixed in a sampling machine known to be practically free from systematic error. They were then withdrawn from the machine one at a time."⁴⁶ The same year, Amberson and his colleagues reported having tossed a coin to decide which of two matched groups of patients would receive gold treatment for tuberculosis.⁴⁷

A few years later, Theobald, a British obstetrician, invited pregnant women to assign themselves at random to a vitamin-supplementation or control group: "An equal number of blue and white beads were placed in a box. Each woman accepted for the experiment was asked to draw a bead from the box. Those who drew blue beads were placed in Group A while those who drew white beads were placed in Group B. The beads drawn out were placed in a separate container."⁴⁸ Theobald acknowledged help from Egon Pearson (Karl's son, and also a statistician) in analysing the results of his study.

Despite these and other examples of random allocation in clinical trials during the 1930s, alternation remained the principal method used to achieve prospective control of biases until well after the end of the Second World War, even in studies by investigators such as Richard Doll, who were very familiar with Fisher's writings.⁴⁹

THE ORIGINS OF FORMAL RANDOMIZATION AND THE PERSISTENCE OF ALTERNATION IN CLINICAL TRIALS

As already noted, some historians believe that the evolution of statistical theory during the first half of the twentieth century helped inspire the randomized clinical trials designed by Austin Bradford Hill and his colleagues in the 1940s. Descriptions of formal randomization in research go back at least a hundred years, however.

Some of the instances noted by others⁵⁰ did not generate comparison groups to evaluate the effects of interventions. In an experiment to assess the ability to distinguish small differences in weights,⁵¹ Pierce and Jastrow wanted to avoid "psychological guessing of what changes the operator (experimenter) was likely to suggest." Initially, they began and ended the series of weights with the heaviest, but then they decided to begin on alternate days with the heaviest and the lightest and used a shuffled deck of cards to decide the order in which to present the different weights to the observers. A few years later, Thorndike and Woodworth⁵² wanted to assess the effects of guessing and correcting estimates of the areas of pieces of paper of varying shapes, so they shuffled these so that observers "could judge their area only from their intrinsic qualities." Neither of these studies applied random allocation to generate comparison groups with a view to making causal inferences about the effects of interventions, but only to keep the assessors unaware. The same applies to the use of playing cards in all but one of the experiments to investigate telepathy reviewed by Ian Hacking;⁵³ in the one exception, John Edgar Coover used dice to decide whether or not a telepathic "agent" should look at a randomly selected playing card before inviting a "reagent" to guess

its identity, thus providing the basis for assessing the effects of the intervention of “looking at the card.”⁵⁴

Early-twentieth-century texts sometimes mentioned the possibility of random allocation in experiments in spheres more comparable to medicine, but there is little evidence that it was adopted in practice. In 1923, William McCall, discussing the design of experiments in education, noted that “equivalence may be secured by chance, provided the number of subjects to be used is sufficiently numerous.”⁵⁵ However, neither in an earlier report co-authored by him,⁵⁶ nor in any other report of an experiment in education in the early twentieth century, have I identified an unambiguous description of random allocation (or alternation, for that matter). Most of the studies to which other writers have referred appear to use matching in attempts to control bias – for example, Winch’s 1908 study to assess the effects of interventions to improve memory in schoolchildren.⁵⁷

Even among statisticians, many of whom, like Fisher, designed agricultural experiments, the origin of randomization remains far from clear. Donald Rubin⁵⁸ has noted: “Despite the early use of physical randomization by Pierce and Jastrow, the allusions to random assignments by ‘Student’⁵⁹ (1923) and the mathematical results using the urn-model formulation in Neyman,⁶⁰ all writers since 1925, including Neyman, seem to agree that the first explicit recommendation to make physical randomization an integral part of experimentation was in Fisher in 1925 and in 1926.⁶¹ This situation, with its juxtaposition of implicit suggestions and explicit contrary attribution from the same author, emphasizes to me the dangers of over interpreting, with ebullient and embellished hindsight, early writings of great men.”⁶²

What is not in any doubt is that Fisher’s 1926 paper – “The Arrangement of Field Experiments” – and his 1935 book – *The Design of Experiments* – affected the design of experiments in agriculture and influenced (and continues to influence) the thinking of theoretical and applied statisticians far beyond that subject.⁶³ I have been unable to find any evidence, however, that Fisher’s writings and conceptualization of the theoretical importance of randomization directly influenced the adoption of randomization in the clinical trials leading up to the MRC streptomycin trial.

Yet despite Fisher’s theoretical considerations, people still used the word “random” to describe alternate allocation to comparison groups in reports of clinical trials during the 1930s and 1940s and in texts on study design, including Bradford Hill’s articles and his *Principles of Medical Statistics*.⁶⁴ If Bradford Hill did not distinguish strict alternation from formal random allocation, this was not because he was

statistically naïve.⁶⁵ His mentors had been Pearson, Yule and, Greenwood, and he used series of random numbers published by members of Karl Pearson's school (who had produced them principally for selecting representative samples from populations).⁶⁶

Although Bradford Hill did not regard himself as a mathematical statistician and had little interest in statistical theory,⁶⁷ he was certainly aware of Fisher's views on the theoretical justification for random allocation. He had known Fisher personally since the 1920s (Fisher had invited him to join the staff at Rothamsted in 1929), and both men held offices in the Royal Statistical Society in the 1930s and 1940s.⁶⁸ Bradford Hill simply did not accord randomization the special status that Fisher and other statistical theorists did. His main interest was in the practical steps required in running clinical experiments, and he adopted randomisation to improve these.

Bradford Hill recognized that the circumstances of experiments in therapeutics were different in important respects from those in agriculture. Thus, in the introductory section of his book, he states: "Elaborate experiments can be planned in which a number of factors can be taken into account statistically at the same time (R.A. Fisher, *The Design of Experiments*, 2nd Edition, 1937, Oliver and Boyd, Edinburgh). It is not my intention to discuss these more difficult methods of planning and analysis; attention is confined to the type of simple experimental arrangement with which medical workers are familiar. Limitation of the discussion to that type must not be taken to mean that it is the best form of experiment in a particular case."⁶⁹ The complex (and statistically efficient) factorial experiments that Fisher had promoted in agriculture were not so readily applicable in clinical research. In agriculture, it is often possible to use already formed study samples – fields, or flocks of sheep, for example. In clinical trials it is usually necessary to assemble study samples over time – for example, patients with pneumonia admitted to hospital over a period of years. The circumstances of clinical medicine are different not only from those in agriculture, but also from those in some other spheres, such as education, where study populations such as school classes exist in their entirety at the beginning of experiments, rather than requiring assembly over time.

An even more important difference between Fisher's and Hill's experiments relates to the essential need for researchers to collaborate with autonomous professionals in clinical research. Hill was aware of the substantial challenge that this presented. Reflecting on it more than half a century later, he wrote: "In (my) articles, I had set out the need for controlled experiments in clinical medicine with groups chosen at random. At the outset, I think I pleaded that trials should be

made using alternate cases. I suspect that if (and it is a very large IF) that were done *strictly*, they would be random. I deliberately left out the words 'randomisation' and 'random sampling numbers' at that time, because I was trying to persuade the doctors to come into controlled trials in the very simplest form and I might have scared them off. I think the concepts of 'randomisation' and 'random sampling numbers' are slightly odd to the layman, or, for that matter, to the lay doctor, when it comes to statistics. I thought it would be better to get doctors to walk first, before I tried to get them to run."⁷⁰

Some commentators have called this retrospective rationalization, but it fits with the testimonies of those who came under Bradford Hill's influence as a teacher. Both Scadding⁷¹ and Doll,⁷² for example, have stated that they used alternation rather than randomization in the studies that they designed during that era probably because of Bradford Hill's exceptionally clear teaching in the 1930s. Bradford Hill believed that medicine could be improved by statistics, and so his starting point was medicine and the medical profession, rather than statistics and theoretical statisticians.⁷³

Bradford Hill's achievement in reaching out to the medical profession was nicely encapsulated in a conversation that he had with John Crofton, who had just presented him for an honorary doctorate in medicine at the University of Edinburgh. Unsurprisingly, Crofton had a good deal to say about Bradford Hill's role in the evolution of controlled trials. A passage in Bradford Hill's unpublished memoir is revealing: "John", I said, 'you know I did not invent the controlled trial. It goes back at least to Lind who tried lime juice in scurvy compared with the usual nauseating mixtures of the day'. 'I know that', Crofton replied, 'but you persuaded an extremely conservative profession which regarded change with suspicion, to accept and use them'. That was, and is, I think a fair judgement," concluded Bradford Hill.⁷⁴

WHY RANDOMIZATION IN THE MRC'S TRIAL OF STREPTOMYCIN?

I concur with Alan Yoshioka's judgment that Bradford Hill's justifications for randomization did not have "much connection with randomisation as discussed in the statistical theory of R.A. Fisher."⁷⁵ So if Bradford Hill's use of "a statistical series based on random sampling numbers" to allocate patients in the streptomycin trial had nothing to do with statistical theory, why was randomization used?

In 1933 Bradford Hill, at the request of the MRC Therapeutic Trials Committee (which had no statistician member), prepared an internal review of an MRC trial of serum treatment for lobar pneumonia.⁷⁶ This

had taken place in four centres and employed a variety of methods, including alternation, to generate control groups.⁷⁷

Bradford Hill's report has disappeared, but Ben Toth⁷⁸ has reconstructed his views of the study from an account that Bradford Hill gave to Stephen Lock⁷⁹ and from Joan Austoker, who read the original report at MRC headquarters in the 1980s.⁸⁰ According to Lock, Bradford Hill criticized the method of allocation in the lobar pneumonia study as insufficiently robust. "He showed, for example, that in the pneumonia trials there were two groups of patients, one of people aged between 20 and 39, and the other aged 40 to 60. Roughly 35 percent of the controls were aged 40 to 60 as opposed to 24 percent of the (serum) patients."⁸¹ Austoker and Bryder reported that Bradford Hill provided detailed criticism of the provision of controls for the trial and recommended greater effort to see "that the division of cases really did ensure a random selection."⁸² Judging by the vast majority of the 51 clinical trials funded under the aegis of the Therapeutic Trials Committee (1931–39), it ignored his advice, and he had little chance of influencing its thinking as he joined it only a year before its demise.⁸³

A report of the serum study appeared a few months after Bradford Hill had submitted his critique of it.⁸⁴ In 1987, Jan Vandenbroucke observed that the published document contains "a beautiful discussion of selection and comparability of treatment groups".⁸⁵ The passage to which he was referring reads as follows:

The good results of insulin on patients with diabetes or of liver treatment in pernicious anaemia are so constant that the trial of these remedies in a very few cases was enough to establish their value. With the antiserum treatment of lobar pneumonia the conditions are very different. The action of the serum is only that of a partial factor for good, and its influence may be overwhelmed by an infection that has been allowed several days to establish its dominance in the patient, or by other complicating factors that weaken the patient's resistance. In order to measure precisely what this partial benefit may be it would be necessary to take two groups of cases of identical severity and initial history and compare the sickness and the fatality in each, the one being treated with serum and the other serving as a control. But this is impracticable, for very few cases, even of "Type 1" lobar pneumonia, are quite alike, and a sufficient number of similar cases could never be got together under one observer and under similar conditions. Some American workers have sought to avoid this difficulty by using a special system of ratings for the various harmful features of the disease, thus expressing each patient's numerical value in reference to a common standard. Such differentiation seemed too intricate, and perhaps too much a matter of personal judgement, for the present inquiry. If a straightforward comparison of treated cases with controls, under the average conditions

whereby patients succeed one another in the wards of a hospital, could not reveal any advantage for those treated by serum, then common sense would conclude that the use of this remedy should be disregarded in the routine of practical medicine. The method consequently agreed upon for London, Edinburgh and Aberdeen was that alternate cases of lobar pneumonia, taken simply in the order of their admission to hospital, should be used respectively for serum treatment and controls. So far as possible both were treated in the same wards and under the care of the same physicians. In the independent inquiry at Glasgow, however, the "serum" cases were treated in the Royal Infirmary, and a series of patients of the same social stratum, admitted during the same period to the Belvedere Isolation Hospital under the care of one physician, served as the control group. It is clear that there may be serious fallacies in any system which contrasts a group of serum-treated patients with a control group drawn from a different stratum of the population, or with a control group in a previous year, when the severity of the prevailing pneumonia might have been different.

Who wrote that paragraph? In 1988, I sent Jan Vandembroucke's published comment to Bradford Hill, who replied: "I feel certain that I wrote that para and I had learned from Pearson & Greenwood & Yule (vide the references No 21 & 22)... I had applied that teaching to the M.R.C's trial of a vaccine against whooping cough and was itching to apply it in the clinical field. Streptomycin provided the opportunity."⁸⁶ Bradford Hill's reference to "that teaching" must presumably refer to alternation, which Pearson had suggested in 1904. Neither alternation nor randomization receives mention in Greenwood and Yule's 1915 article,⁸⁷ nor indeed in the first ten editions of Yule's *Introduction to the Theory of Statistics*. The eleventh edition, co-authored with Kendall and published in the same year as the first edition of Bradford Hill's *Principles of Medical Statistics*, contains sections on "random sampling," but there is no mention of random allocation to comparison groups in intervention studies. However, a section entitled "human bias" opens by declaring: "Experience has, in fact, shown that the human being is an extremely poor instrument for the conduct of a random selection. Wherever there is scope for personal choice or judgement on the part of the observer, bias is almost certain to creep in. Nor is this a quality that can be removed by conscious effort or training. Nearly every human being has, as part of his psychological makeup, a tendency away from true randomness in his choices."⁸⁸

Bradford Hill conceded in his letter to me, "Of course later I may have been influenced by Fisher but not very much – in fact in his famous 'tea and milk' experiment I think he was wrong."⁸⁹ Bradford Hill's statistician son David Hill has explained that his father's objec-

tion "was not to Fisher's analysis of the experiment, but that he thought it would have been a better experiment if the subject had not been told that the 8 cups were to be 4 of each sort."⁹⁰ This comment is consistent with a concern to prevent foreknowledge of allocations among human participants in experiments.

The report of the whooping-cough vaccine trial, which began recruiting a few months before the streptomycin trial, states that the allocation letters A, B, C, and D were "drawn up in random order."⁹¹ Even though the trial was not reported until three years after the streptomycin trial, however, there is no mention of "a statistical series based on random sampling numbers," which might have encouraged the view that considerations of statistical theory lay behind the reference to random allocation. Indeed, Alan Yoshioka⁹² has noted that the word "random" and its derivatives appear nowhere in the MRC files relating either to the whooping-cough trial or to the streptomycin trial! The only explicit reference to Bradford Hill's scheme in the streptomycin trial is a letter referring to "a statistical process of selection," sent by Marc Daniels, the trial's clinical coordinator, to the chair of the steering committee.⁹³

The lack of reference to randomization in the papers relating to the MRC's randomized trial of streptomycin contrasts with proposals for a U.S. randomized trial of streptomycin.⁹⁴ Carroll Palmer, a former member of the Biostatistics Department at Johns Hopkins, was in charge of Public Health Service field studies on tuberculosis and sought controlled studies. Palmer's proposals stated: "The cases chosen by the panel shall, by proper random device, to avoid all possibility of bias, be divided by the Central Unit into cases for treatment and cases for control."⁹⁵

Strict observance of an allocation schedule based on alternation was substantially more probable in a placebo-controlled vaccine trial than in an open trial involving clinical judgments about use of a promising new drug for an often-lethal disease. Thus the *British Medical Journal* (BMJ) noted in its leading article that accompanied the streptomycin report that the panel set up by the trial's steering committee to assess patients' eligibility "conceivably might have been influenced in selecting or rejecting a patient if it had known beforehand whether the patient was to be allocated to the streptomycin or to the controlled group – e.g., if alternate patients had been taken. It was relieved of any such worries by an ingenious system of sealed envelopes. Once a patient had been accepted an appropriate numbered envelope was opened, and not till then was the patient's group revealed. The allocation to "S" or "C" in this form had been made at random by the statistician ... The random allocation has not only removed personal

responsibility from the clinician and possible bias in his process of choosing patients, but has on the whole effectively equated the groups."⁹⁶

Bradford Hill probably wrote the BMJ editorial. There is no relevant documentation, but Stephen Lock, a previous editor of the BMJ who knew Bradford Hill well,⁹⁷ thinks that Bradford Hill told him this. He notes in addition that it is unlikely that the BMJ's pool of editorialists had anybody else who could – or, more cogently, would – have done it, given the quarrelsome temperament of the then editor (who travelled on the same commuter train from home to London as Bradford Hill).⁹⁸

Bradford Hill did not attend the initial meetings of the streptomycin trial's steering committee, but other members had used alternation in clinical trials.⁹⁹ Furthermore, recent testimony from two members of the group – Philip D'Arcy Hart and Guy Scadding – indicates that it adopted the randomization scheme proposed by Bradford Hill not because of statistical considerations, but, as in the patulin trial,¹⁰⁰ because it would help to conceal allocations until after eligible patients had irrevocably entered the study.¹⁰¹ D'Arcy Hart, secretary to the committees overseeing the MRC trials of both patulin and streptomycin, asked in 1996: "Why has [the patulin] trial been overlooked? Is it because attention to the validity of therapeutic trials was generally stimulated by the scheme based on random sampling numbers provided by Bradford Hill to Marc Daniels and me for use in the (streptomycin) trial, which subsequently (from 1948) served as a model for randomization in many later randomized controlled trials? Or is it because the results of the patulin trial were negative and those of the streptomycin trial made medical history?"¹⁰² Certainly the methodological details provided in the report of the patulin trial¹⁰³ and recent oral testimony suggest¹⁰⁴ that care was taken to make the patulin and placebo groups comparable. The report notes: "Previous experience had convinced us that, in a trial of this nature, it is of great importance that both the medical personnel and the patients be prevented from guessing which of the two treatments is genuine and which spurious. It had further been learnt that two solutions are not sufficient to prevent this. In this present trial therefore, four solutions were used, two of which (R and T) contained patulin and two (Q and S) were simply solutions of the buffer salts used in dispensing patulin."¹⁰⁵

In a recent interview, D'Arcy Hart said to me: "Joan [Faulkner], Ruth [Hart] and I went to Cardiff to set up the study. 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 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. Some members of the team that had worked on patulin moved on to do the streptomycin trial."¹⁰⁶

The testimony of D'Arcy Hart and Guy Scadding is consistent with Bradford Hill's response to a more general question, which William Silverman and I put to him when I met him for the first time in 1982. What did he see as the advantages of random allocation over alternation? He made it clear that random allocation was preferable only because it was more likely to prevent advance knowledge of allocations among those involved in recruiting students.¹⁰⁷

BRADFORD HILL ON ALLOCATION, 1937-55

His 1933 analysis of the MRC trial of serum treatment for lobar pneumonia¹⁰⁸ had introduced Bradford Hill to the problems created by clinicians' departures from allocation schemes based on alternation, and he frequently reiterated the need for strict observance of allocation schemes designed to control bias. None the less, Peter Armitage¹⁰⁹ has suggested that Bradford Hill may have continued to underestimate the danger of bias arising from foreknowledge of allocations. Accordingly I trace the evolution of the relevant text in successive editions of *Principles of Medical Statistics* before and after the streptomycin trial.

In the first edition, published in 1937, the relevant passage is the subsection on 'Allocation to groups' in the section on 'The problems of clinical trials' in the concluding chapter. "By the allocation of the patients to the two groups we want to ensure that these two groups are alike except in treatment. It was pointed out in the first chapter that this might be done, with reasonably large numbers, by a random division of the patients; the first being given treatment A, the second being orthodoxly 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. It was also pointed out that this method could be elaborated and the groups made equal in such well defined characteristics as age and sex, and then randomly composed in other respects (and of course, more than one form of treatment could be brought in)."¹¹⁰

There was no change in this wording until the fifth edition (1950), where the final sentence reads: "It was also pointed out that this

method could be elaborated, *or other 'randomising' methods applied* (emphasis added), and the groups made equal in such well defined characteristics as age and sex, and then randomly composed in other respects."¹¹¹

The preface to the sixth edition (1955) signals a substantial change: "In the first edition of this book, issued in 1937, I wrote my final chapter, entitled 'General Summary and Conclusions', round the problem of clinical trials. The discussion was a broad one, of general principles rather than of detail, but, without much change, it appears to me to have stood up reasonably well to the passage of time. That passage of time has, however, brought clinical trials into prominence and fashion, and I have thought it wise in the present edition to take notice of that development. Accordingly, I have introduced a wholly new chapter (Chapter XX) in which the special problems of clinical trials are set out in detail. For use in clinical trials, and many other purposes, I have added 16 pages of random sampling numbers (some 10,000 in all), together with illustrations of how to use them in practice."¹¹²

As far as the pages of random sampling numbers are concerned, David Hill has recorded: "When my father produced his own first set of such tables for his book, he did it by using Tippett to take numbers at random from Kendall and Babbington Smith (or vice versa, I do not know which way round it was) but I remember him saying to me 'If anyone wants to sue me for breach of copyright, they will have to demonstrate which particular digits I have copied.' In later editions I produced pseudorandom numbers for him to use, which I would regard as being, in general, just as good as real randomness, provided that the method is a good one."¹¹³

The subsection on 'The construction of groups' in Bradford Hill's new chapter 20 in 1955 almost certainly reflects the rationale for basing allocation on random numbers in the streptomycin trial.¹¹⁴ The trial used "a statistical series based on random sampling numbers" to ensure 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."

The next step in the setting up of the trial is the allocation of the patients to be included in the treatment and the non-treatment groups (or to more than two groups if more than one treatment is under test). The aim is to allocate them to these "treatment" and "control" groups in such a way that the two *groups* are initially equivalent in all respects relevant to the inquiry. Individuals, it may be noted, are not necessarily equivalent; it is a group reaction that is under study. In many trials this allocation has been successfully made by

putting patients, as they present themselves, alternately into the treatment and control groups. Such a method may, however, be insufficiently random if the admission or non-admission of a case to the trial turns upon a difficult assessment of the patient and if the clinician involved knows whether the patient, if accepted, will pass to the treatment or control group. By such knowledge he may be biased, consciously or unconsciously, in his acceptance or rejection; or through fear of being biased, his judgment may be influenced. The latter can be just as important a source of error as the former but is more often overlooked. For this reason, it is better to avoid the alternating method and to adopt the use of random sampling numbers; in addition, the allocation of the patient to treatment or control should be unknown to the clinician until *after* he has made his decision upon the patient's admission. Thus he can proceed to that decision – admission or rejection – without any fear of bias. One such technique has been for the statistician to provide the clinician with a set of numbered and sealed envelopes. After each patient has been brought into the trial the appropriately numbered envelope is opened (no. 1 for the first patient, no. 2 for the second, and so on) and the group to which the patient is to go, treatment (T) or control (C) is given upon a slip inside. Alternatively a list showing the order to be followed may be prepared in advance, e.g. T, T, C, T, C, C, T, T, T, C, etc., and held confidentially, the clinician in charge being instructed after each admission has been made."¹¹⁵

Two paragraphs about balanced randomization within blocks follow, before the section concludes: "The prescribed random order must, needless to say, be strictly followed or the whole procedure is valueless and the trial breaks down. Faithfully adhered to, it offers three great advantages: (1) it ensures that our personal feelings, or judgments, applied consciously or unconsciously, have not played any part in building up the various treatment groups; from that aspect, therefore, the groups are unbiased; (2) it removes the very real danger, inherent in any allocation which is based upon personal judgments, that believing our judgments may be biased, we endeavour to allow for that bias and in so doing may 'lean over backwards' and thus introduce a lack of balance from the other direction; (3) having used such a random allocation we cannot be accused by critics of having set up personally biased groups for comparison."¹¹⁶

These passages still appeared more than fifteen years later in the ninth edition of the book, with only two sentences added for emphasis at the beginning of the section: "As stated earlier, before admission to a trial *every* patient must be regarded as suitable for *any* of the treatments under study. If this freedom is not present, then equivalent groups cannot be constructed and comparisons are impossible."¹¹⁷ Nowhere does Bradford Hill allude to randomization as a way of

ensuring the validity of tests of statistical significance. His concern continued to be the control of bias, hence his detailed reference to ways of concealing allocation schedules from those involved in recruiting patients for clinical trials.

GRADUAL RECOGNITION OF
THE TWO ESSENTIAL COMPONENTS
OF UNBIASED ALLOCATION IN CLINICAL TRIALS

The 1948 report of the MRC's streptomycin trial is a landmark in the history of clinical trials because of its clear account of how the researchers had implemented the two essential components of unbiased allocation – an unbiased allocation schedule and prevention of foreknowledge of the allocations among those involved in the recruitment of patients. Empirical research today leaves no room for doubt that both of these steps are important,¹¹⁸ but surveys of reports of clinical trials make clear that they remain insufficiently appreciated. This unsatisfactory state of affairs has prompted the emergence of an international initiative by researchers, medical journal editors, and others to improve the situation.¹¹⁹

Given the confusion among clinical researchers, some historians have made incorrect assumptions about the reason that random allocation was adopted for the MRC clinical trial of streptomycin for pulmonary tuberculosis. First, following Fisher, random allocation rather than alternation has been accorded special status in statistical theory. As I have noted, however, not only are the statistical consequences of random allocation and alternate allocation very similar,¹²⁰ but the two methods, and language about them, have intertwined inextricably in the history of clinical trials. Indeed, random allocation has still not replaced alternation, even though schedules based on alternation are substantially more difficult to conceal.

Second, the other element of successful random allocation – concealment of the schedule – may have escaped notice simply because it has had no unambiguous name.¹²¹ Three years after the publication of the sixth edition (1955) of Bradford Hill's book, the statistician David Cox included a section in *The Planning of Experiments* entitled "Randomisation as a device for concealment."¹²² However, only the first of his two examples concerns circumstances "where bias may enter the selection of units to take part in the experiment." His second example is concerned with reducing observer biases in assessing outcomes because randomisation may help to 'blind' observers, whether these are patients, professionals, or researchers. "Allocation concealment" has only recently become a more widely accepted term, distinguishing that particular form of "blinding" from other kinds.¹²³ The latest edi-

tion of Last's *Dictionary of Epidemiology*, for example, defines "Allocation concealment": "A method of generating a sequence that ensures random allocation between two or more arms of a study, without revealing this to either study subjects or researchers. The quality of allocation concealment is enhanced by computer-based random allocation and other procedures to make the process impervious to allocation bias. Less satisfactory methods are allocation by alternation or date of birth, case record, day of the week, presenting or enrolment order."¹²⁴ Even this definition would have been stronger had the final sentence added: "because these permit foreknowledge of the allocations, and thus the temptation to subvert them."

Although I have not found any evidence that statistical theory influenced the MRC's adoption of random allocation for its trial of streptomycin for pulmonary tuberculosis, it is possible that it may have played a more prominent role elsewhere. One of the earliest clear descriptions of formal randomization in a U.S. clinical trial,¹²⁵ for example, reports that a professor of mathematics generated the allocations, and a study published in 1941 used Tippett's random sampling numbers to allocate participants to comparison groups.¹²⁶ Furthermore, textbooks (1938 and 1952) written by one of the first medical statisticians active in North America – Donald Mainland¹²⁷ – pay explicit tribute to Fisher and are far more 'statistical' than Bradford Hill's *Principles of Medical Statistics*. Future research should look at whether statistical theory was more influential in the evolution of clinical trials outside Fisher's home country than within it.

Further research may also show how Bradford Hill's concern to limit bias when controlled trials begin, and Fisher's desire to quantify the uncertainty left after their completion,¹²⁸ merged the 'clinical' and 'statistical' rationales for random allocation and thereby improved the design and analysis of clinical trials during the second half of the twentieth century.

ACKNOWLEDGMENTS

This paper is based on presentations made at the two-day conference "Beating Biases in Therapeutic Research: Historical Perspectives," at the Osler-McGovern Centre, Green College, Oxford, on 4–5 Sept. 2002, which was supported by the UK Cochrane Centre, the Wellcome Unit for the History of Medicine in Oxford, and the Wellcome Trust; and at the three-day conference "La quantification dans les sciences médicales et de la santé: perspective historique et sociologique," at Saint-Julien-en-Beaujolais, France, 24–26 Oct. 2002, which was supported by McGill University, l'Elhess, and the Fondation Mérieux. I am grateful to participants in both these conferences for helpful advice and to

Ben Toth and Alan Yoshioka for copies of their PhD theses. For comments on earlier drafts of this article I am indebted to Doug Altman, Peter Armitage, Luc Berlivet, David Cox, Philip D'Arcy Hart, Richard Doll, David Hill, Michael Kramer, Stephen Lock, Irvine Loudon, Harry Marks, Iain Milne, Keith O'Rourke, William Silverman, Stephen Stigler, Ben Toth, and Ulrich Tröhler. I am especially grateful to Jan Vandenbroucke for his comments on successive drafts and for his studying successive editions of Udny Yule's *Introduction to the Theory of Statistics* and to George Weisz and Annick Opinel for their help and patience in coping with the required format of references.

NOTES

- 1 R. Doll, "The Role of Data Monitoring Committees," in L. Duley and B. Farrell, eds., *Clinical Trials* (London: BMJ Books, 2002), 97–104.
- 2 I. Hacking, *The Taming of Chance* (Cambridge: Cambridge University Press, 1990).
- 3 I. Chalmers, "Comparing Like with Like: Some Historical Milestones in the Evolution of Methods to Create Unbiased Comparison Groups in Therapeutic Experiments," *International Journal of Epidemiology* 30 (2001), 1156–64.
- 4 T.J. Kaptchuk, "Intentional Ignorance: A History of Blind Assessment and Placebo Controls in Medicine," *Bulletin of the History of Medicine* 72 (1998), 389–433.
- 5 H.M. Marks, "Trust and Mistrust in the Marketplace: Statistics and Clinical Research, 1945–1960," *History of Science* 38 (2000), 343–55.
- 6 Medical Research Council (MRC), "Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation," *British Medical Journal* 2 (1948), 769–82.
- 7 D. Mainland, "The Rise of Experimental Statistics and the Problems of a Medical Statistician," *Yale Journal of Biology and Medicine* 27 (1954), 5.
- 8 MRC, "Streptomycin Treatment of Pulmonary Tuberculosis."
- 9 J. Rosser Matthews, *Quantification and the Quest for Medical Certainty* (Princeton, NJ: Princeton University Press, 1995), 127.
- 10 H.M. Marks, *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900–1990*, Cambridge Studies in the History of Medicine (Cambridge: Cambridge University Press, 1997), 132.
- 11 *Ibid.*, 138.
- 12 J.-P. Gaudillière, "Beyond One-Case Statistics: Mathematics, Medicine, and the Management of Health and Disease in the Postwar Era," in U. Bottazzini and A.D. Dalmedico, eds., *Changing Images in Mathematics: From the French Revolution to the New Millennium* (London: Routledge, 2001), 283.
- 13 E. Magnello, "The Introduction of Mathematical Statistics into Medical

- Research: The Roles of Karl Pearson, Major Greenwood and Austin Bradford Hill," in E. Magnello and A. Hardy, eds., *The Road to Medical Statistics* (Amsterdam: Rodopi, 2002), 107.
- 14 Chalmers, "Comparing Like with Like," 1156–64.
 - 15 P. Armitage, cited in *ibid.*, 1157.
 - 16 I. Semmelweis, *Die Aetiologie, der Begriff und die Prophylaxis des Kindbettfiebers* [The Aetiology, Concept, and Prophylaxis of Childbed Fever] (Budapest, 1861); J. Fibiger, "Om Serumbehandling af Difteri" [On Treatment of Diphtheria with Serum], *Hospitalstidende* 6 (1898), 309–25.
 - 17 A. Bradford Hill, *Principles of Medical Statistics*, 1st ed. (London: Lancet, 1937).
 - 18 Armitage, cited in Chalmers, "Comparing Like with Like."
 - 19 See, for example, H.S. Diehl, A.B. Baker and D.W. Cowan, "Cold Vaccines: An Evaluation Based on a Controlled Study," *Journal of the American Medical Association* 111 (1938), 1168–73; I. Sutherland, "Medical Research Council Streptomycin Trial," *Encyclopaedia of Biostatistics* (Chichester: Wiley, 1998), 2559–62.
 - 20 J.F. Box, *R.A. Fisher: The Life of a Scientist* (New York: John Wiley & Sons, 1978), 268–70.
 - 21 R. Kunz, G. Vist, and A.D. Oxman, "Randomisation to Protect against Selection Bias in Healthcare Trials," *Cochrane Methodology Review* 1 (in The Cochrane Library, Oxford: Update Software, 2003).
 - 22 D.G. Altman and K.F. Schulz, "Statistics Notes: Concealing Treatment Allocation in Randomised Trials," *British Medical Journal* 323 (2001), 446–7; K.F. Schulz, I. Chalmers, R.J. Hayes, and D.G. Altman, "Empirical Evidence of Bias: Dimensions of Methodological Quality Associated with Estimates of Treatment Effects in Controlled Trials," *Journal of the American Medical Association* 273 (1995), 408–12.
 - 23 K.F. Schulz, "Subverting Randomization in Controlled Trials," *Journal of the American Medical Association* 274 (1995), 1456–8.
 - 24 A. Bradford Hill, "Discussion on Dr. Armitage's Paper on 'The Comparison of Survival Curves,'" *Journal of the Royal Statistical Society, Series A, Part III* (1959), 295.
 - 25 MRC, "Streptomycin Treatment of Pulmonary Tuberculosis."
 - 26 J. Lind, *A Treatise of the Scurvy. In Three Parts. Containing an Inquiry into the Nature, Causes and Cure, of that Disease, Together with a Critical and Chronological View of What Has Been Published on the Subject* (Edinburgh: Sands, Murray and Cochran for A. Kincaid and A. Donaldson, 1753).
 - 27 W.A. Silverman and I. Chalmers, "Casting and Drawing Lots: A Time Honoured Way of Dealing with Uncertainty and for Ensuring Fairness," *British Medical Journal* 323 (2001), 1467–8, and "Casting and Drawing Lots," James Lind Library (www.jameslindlibrary.org), Thursday 17 April 2003.

- 28 C.H. Parry, "Experiments Relative to the Medical Effects of Turkey Rhubarb, and of the English Rhubarbs, No. 1 and No. II Made on Patients of the Pauper Charity," *Letters and Papers of the Bath Society* 3 (1786), 407–22.
- 29 A.L. Hamilton, *Dissertatio Medica Inauguralis De Synocho Castrensi* (Edinburgh: J. Ballantyne, 1816).
- 30 Ibid.
- 31 I. Milne and I. Chalmers, "Tackling Bias in Assessing the Effects of Health Care Interventions: Early Contributions from James Lind, Alexander Lesassier Hamilton and T. Graham Balfour," *Proceedings of the Royal College of Physicians of Edinburgh* 31, Suppl. 9 (2001), 46–8.
- 32 T.G. Balfour, quoted in *West C. Lectures on the Diseases of Infancy and Childhood* (London: Longman, Brown, Green and Longmans, 1854), 600.
- 33 I. Chalmers and B. Toth, "Thomas Graham Balfour's 1854 Report of a Clinical Trial of Belladonna Given to Prevent Scarlet Fever," James Lind Library (www.jameslindlibrary.org), 17 April, 2003.
- 34 W. Fletcher, "Rice and Beri-Beri: Preliminary Report on an Experiment Conducted in the Kuala Lumpur Insane Asylum," *Lancet* 1 (1907), 1776–9; A. Bingel, "Über Behandlung der Diphtherie mit gewöhnlichem Pferdeserum," *Deutsches Archiv für Klinische Medizin* 125 (1918), 284–332; R.A. Johnston and R.S. Siddall, "Is the Usual Method of Preparing Patients for Delivery Beneficial or Necessary?" *American Journal of Obstetrics and Gynecology* (1922), 4645–50.
- 35 K. Pearson, "Report on Certain Enteric Fever Inoculation Statistics," *British Medical Journal* 3 (1904), 1243–6.
- 36 See www.jameslindlibrary.org
- 37 MRC, "Therapeutic Trials Committee, Serum Treatment of Lobar Pneumonia," *British Medical Journal* 1 (1934), 241–5; W.R. Snodgrass and T. Anderson, "Prontosil in the Treatment of Erysipelas: A Controlled Series of 312 Cases [A Report to the Therapeutic Trials Committee of the Medical Research Council]," *British Medical Journal* 2 (1937), 101–4; W.R. Snodgrass and T. Anderson, "Sulphanilamide in the Treatment of Erysipelas: A Controlled Series of 270 Cases," *British Medical Journal* 2 (1937), 1156–9; T. Anderson, "Sulphanilamide in the Treatment of Measles," *British Medical Journal* 1 (1939), 716–18.
- 38 MRC, "Clinical Trial of Patulin in the Common Cold: Report of the Patulin Clinical Trials Committee," *Lancet* 2 (1944), 373–5.
- 39 P. D'Arcy Hart, "Early Controlled Clinical Trials," *British Medical Journal* (1996), 312.
- 40 See for example Medical Research Council, "Antibiotic and Chemotherapeutic Agents in the Treatment of Infantile Diarrhoea and Vomiting," *Lancet* 2 (1953), 1164–9; MRC, "Thrombophlebitis Following Intravenous Infusions: Trial of Plastic and Red Rubber Giving-Sets," *Lancet* 1 (1957), 595–7.

- 41 S. McDonald, M. Westby, and M. Clarke, "Reporting of the Generation and Concealment of the Allocation Schedule in Randomized Trials in the *BMJ* (1948–1997)," Fourth Symposium on Systematic Reviews: Pushing the Boundaries, Oxford, 2–4 July 2002.
- 42 J.A. Van Helmont, *Oriatrike, or Physick Refined: The Common Errors Therein Refuted and the Whole are Reformed and Rectified* (London: Lodowick-Loyd, 1662), 526.
- 43 Silverman and Chalmers, "Casting and Drawing Lots" (2001), and "Casting and Drawing Lots" (2003).
- 44 D. Colebrook, "Irradiation and Health," Medical Research Council Special Report Series, No. 131 (London: HMSO, 1929).
- 45 J.A. Doull, M. Hardy, J.H. Clark and N.B. Herman, "The Effect of Irradiation with Ultra-violet Light on the Frequency of Attacks of Upper Respiratory Disease (Common Colds)," *American Journal of Hygiene* 13 (1931), 460–77.
- 46 Ibid.
- 47 J.B. Amberson, B.T. McMahon, and M. Pinner, "A Clinical Trial of Sanocrysin in Pulmonary Tuberculosis," *American Review of Tuberculosis* 24 (1931), 401–35.
- 48 G.W. Theobald, "Effect of Calcium and Vitamin A and D on Incidence of Pregnancy Toxaemia," *Lancet* 2 (1937), 1397–9.
- 49 R. Doll, letter to Iain Chalmers, 28 March 2003.
- 50 T. Dehue, "Deception, Efficiency, and Random Groups: Psychology and the Gradual Origination of the Random Group Design," *Isis* 88 (1997), 653–73; A. Oakley, "Experimentation and Social Interventions: A Forgotten but Important History," *British Medical Journal* 317 (1998), 1239–42.
- 51 C.S. Peirce and J. Jastrow, "On Small Differences in Sensation," *Memoirs of the National Academy of Science* 3 (1884), 75–83.
- 52 E.L. Thorndike and R.S. Woodworth, "The Influence of Improvement in One Mental Function Upon the Efficiency of the Other Functions," *Psychological Review* 8 (1901), 247–61.
- 53 I. Hacking, "Telepathy: Origins of Randomisation in Experimental Design," *Isis* 79 (1988), 427–51.
- 54 Ibid., 447.
- 55 W. McCall, *How to Experiment in Education* (New York: Macmillan, 1923), 42.
- 56 E.L. Thorndike, G.J. Ruger, and W.A. McCall, "The Effects of Outside Air and Recirculated Air Upon the Intellectual Achievement and Improvement of School Pupils," *School and Society* 3 (1916), 679–84.
- 57 W.H. Winch, "The Transfer of Improvement in Memory in Schoolchildren," *British Journal of Psychology* (1908), 284–93.
- 58 D.B. Rubin, "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies," *Statistical Science* 5 (1990), 472–80.
- 59 Student, "On Testing Varieties of Cereals," *Biometrika* 15 (1923), 271–93.

- 60 J. Splawa-Neyman, "On the Application of Probability Theory to Agricultural Experiments, Essay on Principles," *Statistical Science* 5 (1990), 465–80. This paper appeared originally in Section 9, *Roczniki Nauk Rolniczych* [Annals of Agricultural Science] 10 (1923), 1–51, and was translated and edited by D.M. Dabrowska and T.P. Speed.
- 61 R.A. Fisher, *Statistical Methods for Research Workers*, 1st ed. (Edinburgh: Oliver and Boyd, 1925); R.A. Fisher, "The Arrangement of Field Experiments," *Journal of the Ministry of Agriculture* 33 (1926), 503–13.
- 62 Rubin, "Comment: Neyman."
- 63 Fisher, "The Arrangement of Field Experiments," and R.A. Fisher, *The Design of Experiments* (London: Oliver and Boyd, 1935).
- 64 A. Bradford Hill, *Principles of Medical Statistics*, 1st ed. (1937).
- 65 Marks, *The Progress of Experiment*.
- 66 D. Hill, letter to Iain Chalmers, 15 April 2003.
- 67 P. Armitage, "Fisher, Bradford Hill and Randomization," *International Journal of Epidemiology* 32 (2003), 925–8; I. Chalmers, "Fisher and Bradford Hill: Theory and Pragmatism?" *International Journal of Epidemiology*, 32 (2003), 922–4.
- 68 Armitage, "Fisher, Bradford Hill and Randomization"; R. Doll, "Fisher and Bradford Hill: Their Personal Impact," *International Journal of Epidemiology* 32 (2003), 929–31.
- 69 A. Bradford Hill, *Principles of Medical Statistics*, 1st ed. (1937), 4.
- 70 A. Bradford Hill, "Memories of the British Streptomycin Trial in Tuberculosis," *Controlled Clinical Trials* 11 (1990), 77–9.
- 71 J.G. Scadding, "Use of Randomisation in Early Clinical Trials," *British Medical Journal* 318 (1999), 1352.
- 72 Doll, "Fisher and Bradford Hill: Their Personal Impact."
- 73 D. Hill, personal communication at meeting on 16 April 2003.
- 74 A. Bradford Hill, "Some Memories of A.B.H.," manuscript dated January 1988, held in the archive at the London School of Hygiene and Tropical Medicine.
- 75 A. Yoshioka, "Use of Randomisation in the Medical Research Council's Clinical Trial of Streptomycin in Pulmonary Tuberculosis in the 1940s," *British Medical Journal* 317 (1998), 1220–3.
- 76 MRC, "Therapeutic Trials Committee."
- 77 I. Chalmers, "MRC Therapeutic Trials Committee's Report on Serum Treatment of Lobar Pneumonia, *British Medical Journal* 1934," James Lind Library (www.jameslindlibrary.org), 12 March 2003.
- 78 B. Toth, "Clinical Trials in British Medicine 1858–1948, with Special Reference to the Development of the Randomised Controlled Trial" (PhD thesis, University of Bristol, 1998).
- 79 S. Lock, "The Randomised Controlled Trial – a British Invention," in G. Lawrence, ed., *Technologies of Modern Medicine* (London: Science Museum 1994), 81–7.

- 80 J. Austoker and L. Bryder, eds., *Historical Perspectives on the Role of the MRC* (Oxford: Oxford University Press, 1989), 46–7.
- 81 Lock, "The Randomised Controlled Trial."
- 82 A. Bradford Hill, "Serum Treatment of Pneumonia, 22 December 1933," cited in J. Austoker and L. Bryder, eds., *Historical Perspectives on the Role of the MRC* (Oxford: Oxford University Press, 1989), 46–7.
- 83 Toth, *Clinical Trials in British Medicine 1858–1948*.
- 84 MRC, "Therapeutic Trials Committee."
- 85 J. Vandembroucke, "A Short Note on the History of the Randomized Controlled Trial," *Journal of Chronic Disease* 40 (1987), 985–7.
- 86 A. Bradford Hill, letter to Iain Chalmers, 7 Aug. 1988.
- 87 M. Greenwood and G.U. Yule, "The Statistics of Anti-Typhoid and Anti-Cholera Inoculations, and the Interpretation of Such Statistics in General," *Proceedings of the Royal Society of Medicine* 8 (1915), 113–94.
- 88 G.U. Yule and M.G. Kendall, *An Introduction to the Theory of Statistics* (London: Charles Griffin, 1937), 337.
- 89 A. Bradford Hill, letter, 1988.
- 90 D. Hill, letter, 2003.
- 91 MRC, "The Prevention of Whooping-Cough by Vaccination," *British Medical Journal* 1 (1951), 1463–71.
- 92 A.Y. Yoshioka, "Streptomycin, 1946: British Central Administration of Supplies of a New Drug of American Origin with Special Reference to Clinical Trials in 'Tuberculosis,'" (PhD thesis, University of London, 1998); A.Y. Yoshioka, "Use of Randomisation in the Medical Research Council's Clinical Trial."
- 93 M. Daniels, letter to G.S. Wilson, 10 Dec. 1946, FD1/6756, National Archives, Kew, London.
- 94 Marks, *The Progress of Experiment*, 122.
- 95 Ibid.
- 96 Leading article, "The Controlled Therapeutic Trial," *British Medical Journal* 2 (1948), 791–2.
- 97 Lock, "The Randomised Controlled Trial."
- 98 S. Lock, e-mail correspondence with Iain Chalmers, 13–21 March 2003.
- 99 MRC, 1944; J.G. Scadding, "Sulphonamides in Bacillary Dysentery," *Lancet* 2 (1945), 549–53.
- 100 MRC, 1944.
- 101 P. D'Arcy Hart, "A Change in Scientific Approach: From Alternation to Randomised Allocation in Clinical Trials in the 1940s," *British Medical Journal* 319 (1999), 572–3; J.G. Scadding, "Use of Randomisation in Early Clinical Trials," *British Medical Journal* 318 (1999), 1352.
- 102 Hart, "Early Controlled Clinical Trials."
- 103 MRC, 1944.
- 104 P. D'Arcy Hart, interview with Iain Chalmers, 2 May 2003.
- 105 MRC, 1944.

- 106 Hart, interview, 2003.
- 107 I. Chalmers, "Why Transition from Alternation to Randomisation in Clinical Trials Was Made," *British Medical Journal* 319 (1999), 1372; Chalmers, "Comparing Like with Like."
- 108 MRC, 1934.
- 109 P. Armitage, "Bradford Hill and the Randomized Controlled Trial," *Pharmaceutical Medicine* 6 (1992), 23-37.
- 110 A. Bradford Hill, *Principles of Medical Statistics*, 1st ed. (1937), 155-60.
- 111 A. Bradford Hill, *Principles of Medical Statistics*, 5th ed. (London: Lancet, 1950), 235.
- 112 A. Bradford Hill, *Principles of Medical Statistics*, 6th ed. (London: Lancet, 1955), v.
- 113 D. Hill, letter, 2003; P. Griffiths and D. Hill, *Applied Statistics Algorithms* (Chichester: Ellis Horwood, 1985), 238-42.
- 114 MRC, 1948.
- 115 A. Bradford Hill, *Principles of Medical Statistics*, 6th ed. (1955), v.
- 116 *Ibid.*, 239-41.
- 117 A. Bradford Hill, *Principles of Medical Statistics*, 9th ed. (London: Lancet, 1971), 254.
- 118 Schulz et al., "Empirical Evidence of Bias," 408-12; Kunz et al., "Randomisation to Protect against Selection Bias."
- 119 D.G. Altman, K.F. Schulz, D. Moher, M. Egger, F. Davidoff, D. Elbourne, P.C. Gøtzsche, and T. Lang, "The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration," *Annals of Internal Medicine* 134 (2001), 663-94 and www.consort-statement.org, 19 April 2003.
- 120 Armitage, cited in Chalmers, "Comparing Like with Like," 1157.
- 121 Schulz et al., "Empirical Evidence of Bias."
- 122 D.R. Cox, *Planning of Experiments* (New York: Wiley, 1958), 79.
- 123 Schulz et al., "Empirical Evidence of Bias"; K.F. Schulz, I. Chalmers and D. Altman, "The Landscape and Lexicon of Blinding," *Annals of Internal Medicine* 136 (2002), 254-9.
- 124 J.M. Last, *A Dictionary of Epidemiology*, 4th ed. (Oxford: Oxford University Press, 2001), 4.
- 125 Doull et al., "The Effect of Irradiation with Ultra-violet Light."
- 126 J.A. Bell, "Pertussis Prophylaxis with Two Doses of Alum-precipitated Vaccine," *Public Health Reports* 56 (1941), 1535-46.
- 127 D. Mainland, *The Treatment of Clinical and Laboratory Data* (Edinburgh: Oliver and Boyd, 1938); D. Mainland, *Elementary Medical Statistics* (Philadelphia: Saunders, 1952).
- 128 H.M. Marks, "Rigorous Uncertainty: Why R.A. Fisher Is Important," *International Journal of Epidemiology* 32 (2003), 932-7.