

# Rigorous uncertainty: why RA Fisher is important

Harry M Marks

Accepted 1 July 2003

Two epistemological claims underwrite the randomized clinical trial (RCT).<sup>1</sup> The first, associated with Austin Bradford Hill, asserts that randomization prevents biased estimates of the value of new therapies. The second, associated with RA Fisher, maintains that randomization is necessary for the valid interpretation of statistical significance.

This paper places Fisher's views on randomization in the larger context of his views on statistical inference and the nature of science. Fisher's life-long interest in the problem of induction, how to draw valid empirical conclusions about the world, led him to emphasize the provisional character of knowledge. The statistical methods he developed—randomization, likelihood—were aimed at producing what Fisher termed 'rigorously specified uncertainty'. Fisher's highly technical arguments about the nature of probability and likelihood are rooted in his more general concerns about the evolutionary and political importance of intellectual autonomy—concerns rooted in his early eugenical views but strongly reinforced by his ideological critique of Soviet science during the Cold War.

Notwithstanding the esoteric origins of some of his arguments, Fisher's ideas have great relevance for contemporary debates within evidence-based medicine. How we best distinguish what is known from what is not known, how we can emphasize the uncertainty of existing knowledge, and how we can best capture that uncertainty—all these Fisherian concerns are crucial in a time when the passage from the journal article to the press conference is increasingly swift.

For various reasons—political, ethical, and practical—the RCT has come under renewed attack in recent years. RCT are said to be too costly and time-consuming; they are unacceptable to some patients and ethically inappropriate in the eyes of some providers. Meanwhile, researchers claim that there are more efficient and equally reliable alternatives for gaining information about medical treatments.<sup>2–5</sup> Such arguments in turn evoke responses which articulate the exceptional value of RCT. At the heart of these defences is the argument that all other methods of evaluating therapy are bias-prone: likely to inflate if not wholly misjudge the value of experimental therapies.<sup>6–11</sup>

As Peter Armitage has suggested, the prevention of bias was central to Austin Bradford Hill's ideas about the randomized

trial.<sup>12,13</sup> Hill and his colleagues introduced randomization in the 1948 Medical Research Council streptomycin trial to prevent physicians from cream-skimming—selectively assigning healthier patients to the experimental drug. For tuberculosis, with its highly variable clinical course, the need for control groups was acknowledged by all. Where assignment to control and treatment groups previously had been handled by alternation, Hill took the process one step further by using randomization, a method which physicians could not so easily manipulate.<sup>14–16</sup>

While randomization was but one among an array of devices introduced into clinical research after World War II to regulate researchers' subjectivity, it was the most novel and possibly the most controversial. Even in the research setting, physicians needed some persuasion to allow the flip of a coin to determine a patient's treatment.<sup>17</sup> Hill accordingly placed great emphasis on randomization's capacity to ensure:

... that neither our personal idiosyncrasies, consciously or unconsciously applied, nor our lack of judgement have entered into the construction of the two (or more) treatment groups and thus biased them in any way...<sup>18</sup>

Nor was Hill alone: a preoccupation with bias and the conviction that randomization, over all other methods of treatment allocation, would prevent bias, was central to most medical presentations of controlled trials in the post-war era.<sup>19</sup>

Hill's arguments for randomization enter into a long historical trajectory of physicians' efforts to provide what Iain Chalmers terms a 'fair test' of innovative therapies.<sup>20</sup> This tradition, Chalmers has argued, dates back to the 18th century therapeutic reformers such as James Lind and John Haygarth whose history has been charted by Ulrich Tröhler.<sup>21</sup> Here I want to explore another venerable tradition in medical statistics—the attempt to quantify, as rigorously as possible, uncertainty about the value of therapeutic (and preventive) practices, using the probability calculus. Like Chalmers' 'fair test' tradition, quantifying uncertainty about medical practices can be tracked back to Daniel Bernoulli's 18th-century efforts to evaluate smallpox inoculation.<sup>22,23</sup> Prior to the 20th century, its intellectual high point was in Jules Gavarret's unsuccessful attempts to apply the probability calculus to therapeutic experiments in mid-19th century Paris.<sup>24</sup> This paper, however, will focus on the statistician, RA Fisher, who introduced randomization into agricultural and scientific experiments but not, as Iain Chalmers has recently emphasized, into medicine.<sup>20</sup>

Institute of the History of Medicine, The Johns Hopkins University, 1900 E Monument St, Baltimore, MD 21205, USA. E-mail: hmarks@jhmi.edu

I have three goals: first, to lay out Fisher's views on randomization, and explain where they fit within his broader notions of statistical inference; second, to place Fisher's ideas about science and statistics in the English intellectual landscape between the Great War and the Cold War; and third, to ask why it was that Fisher's ideas about randomization and uncertainty had so little influence on medical understanding, then or now.

As Armitage suggests, Fisher's ideas about randomization may have been embedded in a more 'theoretically sophisticated' framework than Hill's, but like Hill, Fisher relied on homespun examples to explain the virtues of randomization. Suppose, Fisher argued, an experimenter finds a 10% difference in yields between two grain varieties, planted in different parts of a field. How can you tell whether such a difference reflects a true difference between the varieties, or simply reflects the incidental effects of variations in sunlight, soil fertility, moisture and the like? The standard experimental approach then was to make the two plots as comparable as possible, controlling for all the factors one thought might influence yield. A sometimes useful but insufficient step, Fisher argued. For how does the experimenter **know** that he has captured and successfully controlled for all sources of variability? He cannot.<sup>25,26</sup>

Human beings cannot fully anticipate, Fisher believed, either nature's complexity or her capriciousness. But humans are not helpless, for with randomization, they can produce a rigorous measure of uncertainty. **If** and only if the experimenter randomizes assignments, then he can compare the experimental results with a statistical distribution, and report the probability of a given result (ref. 25, pp. 46–48). As Fisher put it, we should think of experiments as contests with a particularly perverse devil:<sup>27</sup>

To play this game with the greatest chance of success, the experimenter cannot afford to exclude the possibility of any possible arrangement of soil fertilities, and his best strategy is to equalize the chance that any treatment shall fall on any plot by determining it by chance himself. Then if all the plots with a particular treatment have higher yields, it may still be due to the devil's arrangement, but then and only then will the experimenter know how often his chance arrangement will coincide with the devil's.<sup>28</sup>

Where Hill was intensely concerned with controlling bias, Fisher's life-long obsession was with 'rigorously' specifying uncertainty. For Fisher, any statistical inference should indicate 'two things: (1) wherein our factual knowledge differs from complete ignorance and (2) wherein it differs from perfect knowledge' (ref. 27, p. 97). Such inferences could be made from either observational or experimental data, but the data from randomized experiments allowed experimenters to make 'statements of uncertainty' of the 'strongest [most rigorous] possible type'.<sup>29,30</sup>

Statisticians have written a great deal about Fisher's methodological views concerning randomization, probability, and inference.<sup>31,32</sup> Here I want to explore Fisher's philosophical and political ideas about inference. There are arguments against this approach. Fisher was far from being a systematic philosopher and even further from being a member of the political classes. Unlike other 'biometricians', Karl Pearson and Lancelot Hogben, he was not a great popularizer. Most of his work was presented to statistical or scientific audiences, although the

*Eugenics Review* provided a more public outlet for some of his ideas. Still, such an exploration can help us appreciate why the rigorous specification of uncertainty mattered so much to him.

The son of a London auctioneer, Fisher went up on a scholarship to Cambridge, where he was a founding member and intellectual mainstay of the Cambridge Eugenics Society.<sup>33</sup> The British eugenics movement, composed principally of members of the professional middle class like Fisher, was deeply concerned with class differentials in fertility.<sup>34</sup> Fisher's juvenalia offer a clear picture of their concerns: the 'professional middle class', a 'new natural nobility of worth and birth' was being outbred by their biological inferiors, 'the socially lower classes'—a scenario which foretold evolutionary and national disaster if left uncorrected.<sup>35</sup>

The recruits for Fisher's 'new nobility' were to come, not merely from the old intellectual aristocracy, the Darwins and Haldanes, but included those 'skilled artisans' who work with a high degree of intellectual autonomy. 'Versatility'—both intellectual and social adaptability—and the 'mental qualities' to direct one's own work were not, Fisher insisted, a monopoly of the professional middle classes.<sup>36</sup> There is nothing especially unusual about Fisher's efforts to reach across the class line; such efforts were common enough around the time of World War I.<sup>37,38</sup> Fisher's views, however, were noticeably less harsh than those of others in the eugenics movement. What is most interesting here are the echoes of this social (and evolutionary) logic in his later statistical work.

In introducing *The Design of Experiments*, Fisher insisted that the issues discussed:

can be dissociated from all that is strictly technical in the statistician's craft, and, **when so detached**, are questions only of the right use of human reasoning powers, with which all intelligent people, who hope to be intelligible, are equally concerned, and on which the statistician, as such, speaks with no special authority. (ref. 25, p. 2)

Fisher's insistence that any 'thinking man' could understand the 'principles of statistical inference' was not simply rhetorical. Running through *The Design of Experiments* is Fisher's conviction that 'the right use of human reasoning powers' would free 'intelligent people' from a dependence on authority. A proper understanding of statistical inference would allow anyone to say how likely it was that grain A would outgrow grain B, in just the way that it did in a particular experiment. In the absence of a randomized experiment to provide a valid estimate of probability, one was left in the hands of 'scientific authorities', who would pronounce on whether the findings conformed with 'my experience' (ref. 25, pp. 69–70).

Fisher had reason enough to be hostile to experts who threw their weight around, refusing to engage with evidence. At the Population Society, Professor Ernest MacBride had dismissed Fisher's analyses of selection and reproduction as:

of no value, since he is not a trained biologist, and a mathematician's ideas are of no importance in the problem of population.

Fisher had similar run-ins with the physicians on a British Medical Association's committee on sterilization.<sup>39</sup> Yet ultimately

his dislike for an over-dependence on specialists was rooted in his view that too much specialization itself was dysgenic. Versatility and the capacity of intellectual self-governance were the key to evolutionary progress (ref. 35, pp. 314–15).

In his quest for principles of inference that any ‘thinking man’ could understand, Fisher returned, again and again, to the problem of induction; the means by which we gain our knowledge of the natural world. Whereas Fisher considered the development of deductive logic a great accomplishment in human social evolution,<sup>40</sup> he deemed it of limited use in experimental work. Theoretical mathematicians, Fisher argued, lack both the analytical tools and the disposition to deal with uncertainty. They fail to recognize the imperfect nature of empirical knowledge, or to comprehend that a perfectly sound methodology may produce mistaken conclusions.<sup>41</sup>

Like all great inductive philosophers from Hume to Popper, Fisher was a sceptic: ‘Statistical [read ‘empirical’] data are always erroneous in greater or less degree’ (ref. 41, p. 54). But he was an unusually constructive sceptic. Uncertainty and error were, for Fisher, inevitable. But ‘rigorously specified uncertainty’ provided a firm ground for making provisional sense of the world. Randomization was a *sine qua non* of one kind of rigorously specified uncertainty—probability statements. But beyond randomization, a number of stringent requirements must be met before valid probability statements can be made. Fisher’s requirements were that:

- a) there is a measureable reference set (a well-defined set, perhaps of propositions, perhaps of events); b) The subject (of a statement of probability) belongs to the [reference] set; c) No relevant sub-set can be recognized.<sup>42</sup>

Fisher’s various efforts to explain and justify these conditions were and are controversial among statistical theorists. Fisher himself was more sanguine about the limitations of probability statements, for in the many instances where these conditions cannot be met:

a mathematical quantity of a different kind, which I have termed mathematical likelihood, appears to take its place as a measure of rational belief when we are reasoning from the sample to the population. (ref. 41, p. 40; ref. 27, p. 2)

While Fisher believed in the superior rigour of randomized designs, he did not insist on the hard line that emerged in medicine between randomized and non-randomized studies.

For Fisher, both probability statements and likelihood calculations served the same ends: both were rigorously derived statements of uncertainty. But why this emphasis on uncertainty? Again, we return to Fisher’s conviction that empirical knowledge is both highly fallible **and** at the same time corrigible. The point may seem trivially true, and yet for Fisher many statisticians failed to understand its implications. Such failure was at the heart of Fisher’s controversy with Jerzy Neyman over the proper framework for statistical inferences.<sup>43</sup>

In the mid-1930s, Neyman and Egon Pearson elaborated a decision-oriented framework for statistical inference. Inferences, they argued, were like decisions—decisions about what hypothesis to believe. As with any decision (conceived within the framework of modern utility theory), there were loss

functions: defined as the probability of opting for hypothesis A when B was true, or its opposite.<sup>44,45</sup> Fisher had multiple methodological objections to this approach: it placed misguided confidence in probability statements for all inferences; relied overmuch on statistical significance testing; and employed a concept—Type II errors—which Fisher viewed as logically incoherent.<sup>46</sup>

In the early years of their controversy, Fisher’s dissatisfactions were confined to issues of intellectual priority and statistical methodology (ref. 39, pp. 262–66). However, in the Cold War era, Fisher developed a political critique of the ‘ideological’ belief, as he put it, that the world of decisions and the world of science were one and the same.

It was, Fisher argued, entirely appropriate to specify loss functions for incorrect decisions—in the world of corporate, military, or state planning. But in science:

... we introduce no cost functions for faulty judgements. [...] We make no attempt to evaluate these consequences [of inferences] and do not assume that they are capable of evaluation in any sort of currency. (ref. 46, p. 77)

For Fisher, it was crucially important to retain the distinction—and the distance—between the world of science and the decision-making world:

Decision itself must properly be referred to a set of motives, the strength or weakness of which should have no influence whatever on any estimate of probability. We aim, in fact, at methods of inference which should be equally convincing to all rational minds, irrespective of any intentions they may have in utilizing the knowledge inferred. (ref. 46, p. 77)

In eliding the distinction between decisions and inference, Neyman was, Fisher argued, ‘importing from Eastern Europe his misconceptions as to the nature of scientific research’ (ref. 27, p. 186). In the East, decisions about science were a matter of state policy at a time when, Fisher believed, barriers between science and the state were rapidly eroding. By making inferences within a Fisherian framework, statisticians could protect the provisional character of scientific observations, acknowledging the possibility that an inference supported by the data today might turn out tomorrow to be wrong. To Fisher, preserving the distinction between inferences and decisions meant protecting ‘the right of other free minds’ to make ‘their own decisions’, based on a common evaluation of the data (ref. 27, p. 144).

Fisher was hardly alone in articulating such concerns in Cold War Britain: Michael Polanyi and Karl Popper were making related, if more elaborate, arguments about the freedom of science.<sup>47</sup> Those who have puzzled over Fisher’s remark in *Statistical Methods and Scientific Induction*<sup>48</sup> about the need to protect ‘the intellectual freedom that we in the West have taken for granted’ may find the key in his earlier involvement in the Lysenko controversy. TD Lysenko was a Soviet biologist who held heterodox views about the possibility of passing on acquired improvements in biological fitness to subsequent generations. In the late 1940s, he was accused of using his political influence with Stalin to persecute his scientific opponents.<sup>49</sup>

Fisher was among a number of prominent Anglo-American geneticists who had been attacked by the Soviets in the early 1930s for their elaboration of 'bourgeois genetics'.<sup>50</sup> Eighteen years later, rumours that Lysenko had a hand in the death of a scientific critic, Nicolai Vavilov, drew him to the Lysenko debate.<sup>51</sup> In 1948, Fisher, along with JBS Haldane and CD Darlington, did a BBC broadcast about the controversy. Fisher argued that Lysenko was using his political influence with Stalin to intimidate his scientific opponents. Where intimidation was insufficient, Fisher charged, Lysenko helped to see that 'many Russian geneticists' were 'put to death either with or without pre-treatment in a concentration camp.' Fisher devoted the bulk of this inflammatory speech to explaining Lysenko's actions as the product of a system where scientific judgements were made by political leaders.<sup>52</sup>

In 1948, Fisher helped in making Lysenko a stock image of all that was wrong with Soviet science. In 1955, he made use of this caricature in talking about Neyman's statistical ideas at the Royal Statistical Society. After discussing the 'logical differences' between acceptance procedures and scientific inference, Fisher moved quickly to the 'ideological differences':

Russians are made familiar with the ideal that research in pure science can and should be geared to technological performance, in the comprehensive organized effort of a five-year plan for the nation. How far, within such a system, personal and individual inferences from observed facts are permissible we do not know, but it may be safer, and even, in such a political atmosphere, more agreeable, to regard one's scientific work simply as a contributory element in a great machine, and to conceal rather than to advertise the selfish and perhaps heretical aim of understanding for oneself the scientific situation. (ref. 41, pp. 69–70)

What Neyman, a Pole and anti-Bolshevik, had to do with science under the Soviets, Fisher did not explain. In his comments on Fisher's article, Neyman generously stuck to the statistical issues, making no mention of Fisher's vicious efforts at guilt by association.<sup>53</sup>

For someone making heavy use of Cold War rhetoric to indict decision—theoretic approaches, Fisher was surprisingly even-handed in his analysis. In the US, he argued, the danger to independent thought did not come from the state but from the values inculcated by the modern corporation:

In the US also the great importance of organized technology has I think made it easy to confuse the process appropriate for drawing correct conclusions, with those aimed rather at, let us say, speeding production, or saving money. There is therefore something to be gained by at least being able to think of our scientific problems in a language distinct from that of technological efficiency. (ref. 41, p. 70)

By the early 1950s, the place Fisher had imagined for science was threatened, no matter which side of the Iron Curtain he looked towards. In the US, the danger came from the corporation, associated for Fisher with Abraham Wald's decision—theoretical ideas about inference and experimental design; in the Soviet Union, by the long reach of the State, identified with Jerry Neyman's statistical theories.

What, you may be wondering, does this history of Fisher's ideas have to do with randomization and medicine? On first glance, not much. As Iain Chalmers and I have independently argued, Fisher had little to do with the introduction of randomization into medicine. That was the work of Bradford Hill and many others (ref. 17, pp. 148–55).

In the 1950s and 1960s, medical and statistical researchers alike placed great emphasis on randomization's capacity to regulate and control bias (ref. 17, pp. 144–47). Only those who had studied with Fisher or read him carefully discussed randomization's crucial role in ensuring valid estimates of experimental error.<sup>54–56</sup> As a consequence, several generations of physicians grew up ignorant of Fisher's insistence that any experimental result is only an **estimate** of the true difference between treatments. Or that any empirical findings, whether from randomized or non-randomized studies, at best offer **provisional** knowledge of the true situation. In place of these Fisherian notions, a sense of RCT as providing clear-cut 'yes or no' answers emerged in the 1950s and 1960s. In the absence of an established statistical framework for discussing strength of evidence, physicians were uncomfortable with more complex approaches to interpreting experimental findings, especially when the results challenged established medical beliefs (ref. 17, pp. 197–228).

By the 1960s, a growing number of medical statisticians, disturbed by physicians' conviction that 'a significance test is a magic device' to heal all a study's possible ailments, began challenging the importance of 'P-values'; a process of re-education which continues today.<sup>56–59</sup> Similarly, biostatisticians began to re-evaluate the Neyman-inspired doctrine of 'hypothesis-testing', which had helped to reinforce simplistic notions of significance testing.<sup>60</sup> More recently, a renewed theoretical interest in likelihood-based approaches to inference has increased appreciation for 'strength of evidence' concepts and methods.<sup>61</sup> Of greater importance, however, was the growing awareness in the 1980s that on many important clinical issues, there was more than one well-conducted randomized trial with different results. This awareness forced recognition of the need for methods to present, as well to analyse, multiple estimates of a treatment's effects. Much as debate still exists about the appropriate ground rules for meta-analyses, the Cochrane Collaboration's graphic presentation of study findings as point estimates within calculated confidence intervals moves us further away from dogmatic analyses of randomized experiments towards a greater awareness of uncertainty.<sup>58</sup>

RA Fisher saw the development of statistical inference as the 20th century's great contribution to the classical problem of induction: how do we gain reliable empirical knowledge of the world? For Fisher, valid statistical inferences must not only meet certain mathematical conditions; they should also remind us of the provisional character of our knowledge—rigorous uncertainty. Fisher had limited direct influence on the development of the RCT, and even less on the way physicians were taught to understand statistics. Yet in an era where the transit of clinical findings from the *New England Journal of Medicine* to the *Wall Street Journal* grows ever shorter, we are in even greater need of Fisher's reminders about the provisional character of empirical knowledge, and the need to measure uncertainty. I offer no brief for particular Fisherian methods. But some Fisherian attitudes toward statistical inference would do no harm.

## Acknowledgements

I am grateful to Peter Armitage and Iain Chalmers for numerous informative exchanges on the topics discussed in this paper. Peter Hennock, Walter Bodmer, and German Berrios provided helpful leads on British intellectual history; Constantine Frangakis and Richard Royall did the same on questions of statistical theory. Shah Ebrahim provided valuable advice on rewrites. None of the above is responsible for the author's failure to follow their tutelage.

## References

- 1 Rosenberger WF, Lachin J. *Randomization in Clinical Trials. Theory and Practice*. New York: John Wiley, 2002.
- 2 Abel U, Koch A. The role of randomization in clinical studies: myths and beliefs. *J Clin Epidemiol* 1999;**52**:487–97.
- 3 Horton R. The clinical trial: deceitful, disputable, unbelievable, unhelpful, and shameful—what next? *Control Clin Trials* 2001;**22**: 593–604.
- 4 Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;**342**:1878–86.
- 5 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies and the hierarchy of research designs. *N Engl J Med* 2000;**342**:1887–92.
- 6 Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 2000;**342**:1907–09.
- 7 Gilbert JP, McPeck B, Mosteller F. Progress in surgery and anesthesia: Benefits and risks of innovative therapy. In: Bunker JP, Barnes BA, Mosteller F. *Costs, Risks and Benefits of Surgery*. Oxford: Oxford University Press, 1977, pp. 24–169.
- 8 Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: Medical. *Stat Med* 1989;**8**:441–54.
- 9 Miller JN, Colditz JA, Mosteller F. How study design affects outcomes in comparisons of therapy. II: Surgery. *Stat Med* 1989;**8**:455–66.
- 10 Ioannidis JPA, Haidich AB, Pappa M *et al*. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;**286**:821–30.
- 11 Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;**317**:1185–90.
- 12 Armitage P, Bradford Hill and the randomized controlled trial. *Pharmaceutical Medicine* 1992;**6**:23–37.
- 13 Armitage P, Fisher, Bradford Hill, and Randomization. *Int J Epidemiol* 2003;**32**:925–28.
- 14 Hart PD. A change in scientific approach: from alternation to randomised allocation in the clinical trials in the 1940s. *BMJ* 1999;**319**:572–73.
- 15 Yoshioka A. Use of randomisation in the Medical Research Council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940s. *BMJ* 1998;**317**:1220–23.
- 16 Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002;**359**:614–18.
- 17 Marks HM. *The Progress of Experiment. Science and Therapeutic Reform in the United States, 1900–1990*. New York: Cambridge University Press, New York, 1997, pp. 155–58.
- 18 Hill AB. Assessment of therapeutic trials. *Trans Med Soc London* 1953;**68**:128–47.
- 19 Marks HM. Trust and mistrust in the marketplace: statistics and clinical research, 1945–1960. *History of Science* 2000;**38**:343–55.
- 20 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.
- 21 Tröhler U. *To Improve the Practice of Medicine. The 18th Century British Origins of a Critical Approach*. Edinburgh: Royal College of Physicians of Edinburgh, 2000.
- 22 Daston LJ. *Classical Probability in the Enlightenment*. Princeton: Princeton University Press, 1988, pp. 82–89.
- 23 Marks HM. *Should the State Count Lives? An Eighteenth-Century Controversy Over Smallpox Inoculation*. Conference on Quantification dans les Sciences Médicales, Fondation Mérieux, 23–25 October 2002.
- 24 Matthews JR. *Quantification and the Quest for Medical Certainty*. Princeton: Princeton University Press, 1995, pp. 14–85.
- 25 Fisher RA. *The Design of Experiments*. London: Oliver & Boyd, 1935, pp. 35–38, 46–49, 68–70.
- 26 Fisher RA. The arrangement of field experiments. *J Ministry of Agriculture of Great Britain* 1926;**33**:503–13.
- 27 Bennett JH (ed.). *Statistical Inference and Analysis. Selected Correspondence of R.A. Fisher*. Oxford: Clarendon Press, 1990, p. 269.
- 28 Box J. RA Fisher and the design of experiments, 1922–1926. *Am Stat* 1980;**34**:1–7.
- 29 Fisher RA. *The Place of the Design of Experiments in the Logic of Scientific Inference*. Paris: Le Plan de Experiences, Editions de la Recherche Scientifique, 1963, p. 15.
- 30 Fisher RA. The underworld of Probability. *Sankhya* 1957;**18**:201–10.
- 31 Savage LJ. On rereading RA Fisher. *Ann Stat* 1976;**4**:441–500.
- 32 Aldrich J. RA Fisher and the making of maximum likelihood 1912–1922. *Stat Sci* 1997;**12**:162–76.
- 33 Mazumdar PMH. *Eugenics, Human Genetics and Human Failings. The Eugenics Society, Its Sources and Critics in Britain*. London: Routledge, 1992, pp. 96–103.
- 34 Soloway RA. *Demography and Degeneration. Eugenics and the Declining Birthrate in Twentieth-Century Britain*. Chapel Hill, NC: University of North Carolina Press, 1990.
- 35 Fisher RA. Some hopes of a eugenicist. *Eugenics Review* 1914;**5**: 309–15.
- 36 Fisher RA. Positive eugenics. *Eugenics Review* 1917;**9**:206–12.
- 37 Perkin H. *The Rise of Professional Society. England Since 1880*. London: Routledge, 1990.
- 38 McKibben R. *Classes and Cultures. England 1918–1951*. London: Oxford University Press, 1998, pp. 50–59, 98–102.
- 39 Box J. RA Fisher. *The Life of A Scientist*. New York: John Wiley & Sons, 1978, pp. 194, 198–99.
- 40 Fisher RA. The bearing of genetics on theories of evolution. *Science Progress* 1932;**27**:273–87.
- 41 Fisher RA. The logic of inductive inference. *J R Statist Soc* 1935;**98**:39–82.
- 42 Fisher RA. The nature of probability. *Centennial Review of Arts and Sciences* 1956;**2**:261–74.
- 43 Bartlett, MS. Jerzy Neyman 16 April 1894–5 August 1981. II. Neyman and the theory of statistical inference. *Biog Mem Fellows R Soc* 1982;**28**:390–98.
- 44 Neyman J, Pearson ES. On the problem of the most efficient tests of statistical hypotheses. *Philos Trans R Soc London A* 1933;**321**: 289–337.
- 45 Neyman J. Basic ideas and some recent results of the theory of testing statistical hypotheses. *J R Statist Soc* 1942;**105**:292–327.
- 46 Fisher R. Statistical methods and scientific induction. *J R Statist Soc B* 1955;**17**:69–78.
- 47 McGucken W. *Scientists, Society and State. The Social Relations of Science Movement in Great Britain 1931–1947*. Columbus, OH: Ohio State University Press, 1984, pp. 275–91.
- 48 Fisher RA. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver & Boyd, 1956, p. 7.

- <sup>49</sup> Kremontsov N. *Stalinist Science*. Princeton: Princeton University Press, 1997, pp. 159–83.
- <sup>50</sup> Jones G. *Science, Politics and the Cold War*. London: Routledge, 1988, p. 19.
- <sup>51</sup> Graham LR. *Science in Russia and the Soviet Union. A Short History*. New York: Cambridge University Press, 1993, pp. 128–31.
- <sup>52</sup> Fisher RA. The Lysenko controversy. *The Listener* 1948;**40**:874–75.
- <sup>53</sup> Neyman J. Note on an article by Sir Ronald Fisher. *J R Statist Soc B* 1956;**18**:288–94.
- <sup>54</sup> Mainland D. Statistics in clinical research: some general principles. *Ann N Y Acad Sci* 1950;**52**:922–30.
- <sup>55</sup> Greenberg BG. Why randomize? *Biometrics* 1951;**7**:309–22.
- <sup>56</sup> Mainland D. The use and misuse of statistics in medical publications. *Clin Pharmacol Ther* 1960;**1**:411–22.
- <sup>57</sup> Goodman SN, Royall R. Evidence and scientific research. *Am J Public Health* 1988;**78**:1568–74.
- <sup>58</sup> Altman DG. Statistics in medical journals: some recent trends. *Stat Med* 2000;**19**:3281–84.
- <sup>59</sup> Sterne J, Davey Smith G. Sifting the evidence: what's wrong with significance tests. *BMJ* 2001;**322**:226–31.
- <sup>60</sup> Cutler SJ, Greenhouse SW, Cornfield J *et al*. The role of hypothesis testing in clinical trials. *J Chronic Dis* 1966;**19**:857–82.
- <sup>61</sup> Royall R. On the probability of observing misleading statistical evidence. *J Am Statist Assoc* 2000;**95**:760–80.