

Mainland's *Elementary Medical Statistics* (1952): a pivotal text in statistical pedagogy

Vern Farewell 

MRC Biostatistics Unit, University of Cambridge, UK

Corresponding author: Vern Farewell. Email: vern.farewell@mrc-bsu.cam.ac.uk

Introduction

Austin Bradford Hill, following in the footsteps of Major Greenwood, is regarded as the leading figure in the development of medical statistics in the United Kingdom during the middle years of the 20th Century. His influence also extended beyond the United Kingdom through his publications and international collaborations. In the United States, Raymond Pearl was an early proponent of the importance of medical statistics, and a contemporary of Greenwood. Also, individuals such as Harold Dorn, William Cochran and Donald Mainland held senior posts in medical statistics or public health contemporaneously with Hill.

More specifically, and although the story of the development of clinical trials is not straightforward,¹ Hill is sometimes characterised as ‘the father of the modern clinical trial’ because of his early use of randomisation in trials, notably the well-known MRC streptomycin trial.² Additionally, in epidemiology, the Bradford Hill Criteria for Causation have been widely used for exploring issues of causation in non-randomised epidemiological studies for many years. Both topics feature in a 1952 textbook, *Elementary Medical Statistics*,³ by Mainland, who at the time of publication was Professor of Medical Statistics in the Department of Preventive Medicine at New York University.

The purpose of this article is to highlight the contributions of Mainland's text and also, and more specifically, to suggest why his writing reflects, much more strongly than Hill's, the statistical thinking of Sir Ronald Fisher, who developed many of the methods for the design and analysis of experiments that dominated the statistical landscape during the period when Mainland and Hill were writing. With regard to the specific context of treatment allocation, which is discussed in this article, this has also been addressed by Matthews.^{4–6}

Historical background to the state of medical statistics in 1952

In the early years of the 20th century, the most influential medical statistician in the United Kingdom was Major

Greenwood.⁷ Greenwood played a major role in the early days of the UK Medical Research Council (initially Committee), chairing a Medical Committee that worked in tandem with the MRC Statistical Department headed up by the medically qualified John Brownlee⁸ until his early death in 1927. After Brownlee's death, MRC statistical activities were consolidated under Greenwood's leadership with the Statistical Department moving to the London School of Hygiene and Tropical Medicine (LSHTM) where Greenwood had been appointed Professor.

When Greenwood, who was also medically qualified, moved into research activities, he sought out training and advice from Karl Pearson at University College London. While Pearson was primarily interested in what was then termed biometry, which largely focused on biological applications, Greenwood recognised that Pearson's methods would be valuable in medical research. Furthermore, John Brownlee was also known as a ‘disciple’ of Pearson although he did not, apparently, have any contact with Pearson beyond his publications.⁸

As a result, in the United Kingdom, early work in what might be termed modern medical statistics, in contrast to previous work which was largely demographic and related to official statistics, was primarily informed by Pearsonian methods. Concurrently however, R.A. Fisher was developing the foundations of modern statistics, in general, in the context of experimental work at the Rothamsted agricultural research establishment. Generally, and specifically through his involvement in the Royal Statistical Society, Greenwood would have known of Fisher's work, but his background probably made it unlikely that he would engage with the technical details. It can be noted, however, that Greenwood recognised the need for individuals with mathematical skills in statistical work, such as Isserlis and Newbold,⁹ and used mathematical developments in some of his own research, even in his last article on the topic of accidents.¹⁰

Greenwood's protégé, who would succeed him at the LSHTM and in directing MRC statistical activity until the 1960s, was Austin Bradford Hill. Hill became involved in medical research under Greenwood's direction after being

discharged for medical reasons out of World War I and taking a correspondence degree in Economics at LSE while convalescing. As Armitage records,^{11–13} Hill was not interested in the theoretical aspects of statistical research but was primarily concerned with seeing statistics properly used in medical research. To this end, Hill wrote a series of articles for *The Lancet* that formed the basis of his famous book, *Principles of Medical Statistics*,¹⁴ first published in 1937 and with 11 further editions, the last in 1991. Hill and Fisher knew each other,¹³ but Fisher's statistical methods are not mentioned in any detail in Hill's writing. This is the case in spite of there being some suggestion of Fisher's influence on Hill's colleagues Woods and Russell¹⁵ and, very directly, on Oscar Irwin¹² who was in Hill's department and had studied with Fisher. Nevertheless, Fisher was complimentary of Hill's book,¹⁶ although it did not reflect major themes of Fisher's writings. A more comprehensive assessment of Hill's attitude to Fisher's work is given by Matthews.⁵

Hill was internationally recognised as a leader in medical statistics and visited the United States on a number of occasions. However, the development of medical statistics in the United States followed a somewhat different course than that in the United Kingdom. While not providing here a comprehensive assessment of this development, an apparent difference relating to the influence of Fisher can be noted. Raymond Pearl was an early advocate of statistical methods in medical research in the United States and he, like Greenwood of whom he was a contemporary, also studied with Karl Pearson.¹⁷ And, Pearl's book, *Medical Biometry and Statistics*¹⁸ makes very little reference to Fisher, perhaps reflecting the antipathy between Pearson and Fisher at that time. This is also true in the third edition published in 1940,¹⁹ although that edition does credit Fisher with determining the appropriate degrees of freedom for the chi-squared test, without mentioning that Fisher's arguments corrected those of Pearson.

In the 1950s however, the US situation changed. After undergraduate study at Cambridge, William Cochran (see <https://mathshistory.st-andrews.ac.uk/Biographies/Cochran/>) had started a PhD there but was convinced that a post with Frank Yates at Rothamsted would be better, even though his Cambridge PhD time had led to his first article²⁰ outlining what is now known as Cochran's theorem, the theoretical basis for the *F*-tests associated with analysis of variance. However, in 1939, Cochran left the United Kingdom for a post at the Iowa Statistical Laboratory where the American statistician George Snedecor was Director. Snedecor recognised the value of Fisher's work early in his career, and Fisher visited Iowa State in the summers of 1931 and 1936.²¹ Cochran's extensive exposure to Fisherian statistics was brought into medical statistics when he accepted, in 1948, the Chair of the Johns Hopkins Department of Biostatistics, a post he held

for 10 years. Notably, while there, he was involved in the trial of the polio vaccine, introducing some randomisation and arguing (unsuccessfully) to allow for over-dispersion due to clustering.²² Around the same time, in 1950, Mainland, who had spent summers with Fisher in the 1930s, moved from Canada to the United States to take up a post as Professor of Medical Statistics in New York University. As will be seen, Mainland's book is imbued with Fisher's statistical arguments. A third leading figure, Harold Dorn,²³ joined the US National Institutes of Health in 1948 and until 1961 was 'considered, *de facto* if not *de jure*, chief statistician of the NIH'.²⁴ Harold Dorn, although his academic genealogy connected to Pearson at University College London, was, by training, a sociologist interested in surveys. However, he worked with Fisher, and Egon Pearson, when he spent the academic year 1933–1934 at University College London. Fisher, who had no specific interest in medical statistics himself (outside of genetics), was suddenly, therefore, very much an influence on medical statistics in the United States.

In general, it can be conjectured that, at this time, there was also a shift in the approach to medical statistics as mathematically well-trained statisticians, who would be familiar with statistical theory, began to move into the field. Notable examples were Jerry Cornfield in the United States who joined Dorn at the NIH and Peter Armitage who joined Hill at the LSHTM. However, it would take a few years before the influence of such appointments would be felt, and this would certainly have been after the publication of Mainland's 1952 book. In assessing the role of Mainland's book therefore, it is the comparative experience and interests of the senior medical statisticians at that time that is most relevant.

Donald mainland and elementary medical statistics

Doug Altman²⁵ has provided an excellent biographical article on Mainland and a basic summary of his career is given in the opening paragraph of Altman's 'Brief Biography' section: 'Donald Mainland graduated in medicine at Edinburgh. He taught anatomy in Edinburgh and received a Doctor of Science degree there for his research in embryology and histology. In 1927, he moved to Winnipeg, Manitoba, Canada, and in 1930, at the age of 28, became Professor and Chairman of the Department of Anatomy at Dalhousie University. Even his earliest publications showed an interest in measurement issues, and foreshadowed an increasing interest in statistics. In 1938, he published his first book on statistics in medicine. In 1950, he became Professor of Medical Statistics at the New York University and shortly afterwards published his best-known book, *Elementary Medical Statistics*. Thereafter, Mainland was a prolific and influential writer on statistical topics.'

Mainland's initial research work was largely laboratory-based, for example on the embryology of ferrets. However, a colleague in Winnipeg introduced him, in 1928, to Fisher's book, *Statistical Methods for Research Workers*, first published in 1925.²⁶ Mainland quickly recognised the relevance and importance of statistical ideas to his own work. His interest grew and in 1934, after presumed correspondence with Fisher, he was invited by Fisher in April 1934 to visit Fisher that summer in London. There were additional summer visits in the late 1930s. It seems likely therefore that Mainland would have discussed with Fisher not only the content of *Statistical Methods for Research Workers* but also Fisher's *Design of Experiments*.²⁷

Mainland's first book on statistics in medicine, published in 1938,²⁸ focused primarily on laboratory data and, as Altman notes, 'Fisher was thanked profusely in the preface'. More general coverage was provided in a 1948, 166-page, journal article, 'Statistical methods in medical research. I qualitative statistics (enumeration data)'.²⁹ This was towards the end of his time at Dalhousie in Halifax where he taught courses on statistics. These courses provided the basic material for his 1952 book, although it was published after he had moved to New York University. Parenthetically, Part II,³⁰ dealing with sample sizes, appeared 5 years later and was only 11 pages in length.

Why did Mainland write *Elementary Medical Statistics*? One answer to this question is found in the first paragraph of the book's section entitled 'The Purpose of This Book'. It reads

'Because the neglect of statistical methods in medicine is due largely to faulty training of students, some medical schools are now trying to correct the fault, and this book is an outgrowth of such attempts. It is designed primary for students who are to become practitioners; but the principles and techniques are the same as those needed by investigators at the beginning of their careers, for they are common to all branches of medicine, from histology to psychiatry'.

But why did Mainland write a book with this purpose? There are perhaps three reasons. The first is that Mainland had for many years been a medical researcher but had early on been convinced of the importance of statistical thinking. As Chairman of the Department of Anatomy in the Dalhousie Medical School, he published a major 1938 textbook on anatomy but he also developed courses in statistics, first for students of anatomy and then for the medical school more generally. As Altman documents, however, Mainland's move to a position as Professor of Medical Statistics was made so that he could concentrate on statistics, and have additional scope to become involved in clinical trials. Although he was a self-taught statistician, statistics was clearly now his primary interest. The second reason was that his road to an interest in

statistics made him wary of the obvious solution to the 'neglect of statistical methods in medicine' which was to make statistics part of the medical curriculum. As he writes in the first sentence of his book's preface, 'Those who have for many years stressed the importance of statistical thinking in medicine cannot be entirely happy to see statistics become established as a subject in the undergraduate curriculum . . .'. This sounds counter-intuitive or self-contradictory, but his worry is that this will shift the focus to 'board examinations which foster static pedagogy' and undermine the potential for statistics in medicine to break down 'interdepartmental barriers' and to establish 'a set of principles by which we can draw valid conclusions from experience'. The third and pragmatic reason is that he had developed course notes, and this would provide a broader outlet for their use while propagating his views on medical statistics.

His views are reflected in the style of the book. Although he gives, for example, the details of how to perform chi-squared tests on categorical (described by Mainland as 'enumeration') data and *t*-tests on continuous ('measurement') data, he intersperses this with broad ranging discussions of when they should be used and how they should be interpreted. This intermingling in presentation is intentional, and important enough to him that he adopts it. This is clear because he signposts in the introductions of various chapters that a 'student' might better study the pages dealing with the details of the methods before returning to the material which relates to broader issues regarding their application.

Some of these broader issues will be highlighted in the Appendix for this article, which examines the chapters of Mainland's book individually. However, the issue of randomisation, and causal inference more generally, is of particular interest at the present time to statisticians and medical researchers more generally, and the coverage of this topic in *Elementary Medical Statistics* is examined separately in the next section.

Randomisation and causal inference

Randomisation *per se* is first mentioned in Chapter 2, titled 'On Looking at Evidence', under the heading 'Lack of Objectivity' (p. 27). Mainland uses the 1948 publication reporting the results of the MRC streptomycin trial in pulmonary tuberculosis to illustrate how objectivity must be planned into an investigation. His first observation is that, prior to the trial, there were indications that streptomycin might be more beneficial than alternative treatments and that, therefore, physicians might want to preferentially give this treatment to certain patients. Mainland, at this point, says simply 'This risk was avoided by the sampling method that will be described in Chapter 4'. That method is randomisation. A key feature

of this procedure is likely to be independence of assignment. His second comment on objectivity relates to blinded reading of x-ray films, which was also a feature of the MRC trial.

It is in the section 'On Planning a Simple Experiment' in Chapter 4 that the term 'Random Sampling' appears as a sub-heading. In fact, 'random' is a very broad word but, in the context of experimentation, it can be regarded as arising when experimental subjects, or items, are assigned a treatment by an 'objective impersonal procedure'.³¹ In this section on random sampling, after discussing the value of systematically ensuring a balance between two treatment groups with respect to some known risk factors in reducing variability, Mainland writes (p. 103) that this should be done 'not as far as possible, but *as far as it is convenient and useful* [his emphasis]'. To deal with the inevitable differences in other factors that might influence outcome, which cannot be distributed equally to the two treatments, he writes 'we must allocate them in such a way that we can tell what allowance to make for the inequalities'. Furthermore, 'The only way to do this is to make chance decide for us'.

In these two excerpts, Mainland is addressing, respectively, the two aspects of randomisation that are outlined by David Cox in his book *Planning of Experiments* (p. 85).³¹ These are:

- 'that in a large experiment, it is very unlikely that the estimated treatment effect will be appreciably in error'
- 'that the random error of the estimated treatment effects can be measured and their level of statistical significance examined, taking into account all possible forms of uncontrolled variation subject to (1)' where (1) refers to the error structure assumed.

The first aspect relates to allowing the definition of an unbiased *estimator* of a defined treatment difference, the *estimand* of interest. Generally, a key feature of this is the avoidance of confounding. The second relates to the properties of the *estimation* process.

Alternate allocation

In 1952, in clinical trials, the first purpose of randomisation was increasingly recognised, but there was frequent reference to a supposed method of randomisation based on alternate allocation of patients to the two treatments under study. Matthews⁴⁻⁶ has provided a careful investigation of the history of this method and highlights Mainland's criticisms of this methodology. In the section in Chapter 4 on random sampling in *Elementary Medical Statistics*, Mainland simply highlights that alternate allocation 'is not strictly equivalent to random sampling' and gives some possible examples of systematic differences

that might be introduced. He returns to this topic in a later chapter (p. 268): here he gives a further example of how an unknown 'rhythm', due to differential risks and patient numbers on days of the week, could bias results. He concludes by saying 'When an investigator employs a method that is not strictly random as if it were equivalent to a random technique, the onus is on him to prove it justifiable by experiment, not argument; and this, even if possible, would entail a very large investigation'.

These arguments reflect Mainland's belief, outlined in the section 'How Randomization Acts' in Chapter 4 (p. 104), that 'Random sampling is the only way to *equalize the risk of hidden bias*' [his emphasis].

As Matthews⁶ highlights, Mainland's strong aversion to alternate allocation, reflecting Fisher's influence, was not shared by other medical statisticians, notably Hill, at least during this time period. Armitage writes, about Hill's later (1990) reflection on the MRC streptomycin trial, that Hill felt 'that alternations of successive cases might have been successful if strictly adhered to, but he (Hill) wrote "it's a very big IF"'. However, Matthews very appropriately highlights that Hill's writings largely do not reflect this perspective, and he often presents it as a credible method of treatment allocation.

Mainland's distaste for an approach that carries inferential risks is more generally expressed in his otherwise favourable review³² of the book *Controlled Clinical Trials*³³ which is a collection of articles, edited by Hill, given at a 1959 conference on clinical trials in Vienna. While recommending the book, Mainland did introduce one caution. He suggested, referring to a comment by Hill, that the reader should not be 'beguiled by a statistician's understatement (p. 17) that "however carefully a trial has been planned occasionally things will go wrong"'. He then highlights a clinician's contribution (p. 166) that 'a clinical trial is a serious matter and is not to be undertaken lightly . . . It is important to get the answer right, and this means that an enormous amount of trouble has to be taken in the planning, execution and publication of a trial'. This is clearly Mainland's view.

Error estimation

Mainland's insistence on randomisation, or experimental justification of any alternative, could be said to reflect the high priority Fisher gives to randomisation in his writing. However, in the context of agricultural experiments in which Fisher primarily worked, Fisher seems to simply assume the value of randomisation to prevent biased treatment estimates and focuses in much more detail on the issue of error estimation. For example, in his book, *Design of Experiments*²⁷ on page 72, Fisher's section on 'Bias in Systematic Arrangements' illustrates how systematic arrangements thought to bring balance to treatment assignment can lead to incorrect estimates of error

for the treatment comparison. The error variance is larger than it should be when the systematic arrangement serves to eliminate the variation due to a particular risk factor, but the analysis then assumes that the treatment assignment is random over this factor. If the systematic arrangement happens, incorrectly for whatever reason, to increase imbalance with respect to the risk factor, then the error variance can be artificially reduced. Thus, the control of confounding and assumptions about the error structure themselves become confounded. This discussion reflects Fisher's earlier arguments, pages 20–24, that randomisation provides a physical basis for the validity of significance tests concluding 'the simple precaution of randomisation will suffice to guarantee the validity of the test of significance, by which the result of the experiment is to be judged'.

Mainland certainly affirms this value of randomisation. Immediately following his section 'How Randomization Acts', Mainland addresses 'Interpretation after Random Sampling'. The point he makes here is that the additional benefit of randomisation is that '...we can, after the experiment, use our knowledge of chance to interpret the results'. This clearly relates to the second aspect of randomisation. However, Mainland does perhaps pay more attention to the possibility of bias from unknown factors. This is addressed in the section 'How Randomization Acts' as discussed earlier, and he again addresses it in the section 'Interpretation after Random Sampling' where he addresses the possibility of 'something else' influencing a significance test result. He writes, 'If treatments V and W were not allocated at random, the something else might be ... some bias due to unknown factors'. In addition, Mainland's illustrations of the problems with non-randomised studies, discussed in the previous section, relate to systematic arrangements leading to bias in the estimation of the treatment effect, not, or not solely, in the estimation of the error variance of that effect.

But, and it is an important but, Mainland's recognition of the use of randomisation to justify tests of significance contrasts sharply with the writings of Hill, who, as far as it appears, pays little or no attention to this matter. As Armitage³⁴ says, and Matthews,⁵ following Silverman and Chalmers,³⁵ highlights, Hill had little interest in statistical theory in general. Indeed, Matthews goes further to argue that (a) Hill's attitude was not just indifference to 'technicalities' but was actively dismissive and (b) this led to his failure to see the importance of randomisation in drawing inferentially reliable conclusions from clinical trials.

Alternate allocation and error estimation

In a general sense, Armitage (2003)¹³ argued that Hill's views of randomisation were influenced by the context in which he advocated it. Basic principles of comparative

experimentation were widely accepted in the agricultural setting within which Fisher worked, but fundamental concepts such as the need for simultaneous control were still having to be emphasised in the medical context. Alternate allocation might have been seen as the simplest thing to advocate for this control to those for whom randomisation might have seemed difficult to implement or somehow incompatible with medical care.^{4–6}

Mainland's insistence on randomisation in clinical trials may have been, as Matthews argued, primarily motivated by his understanding of the broader arguments for randomisation, and, perhaps, because his original medical research was in the context of laboratory studies in anatomy, within which the introduction of randomisation would have been less problematic than in trials. Combined with his personal exposure to Fisher, his prescient advocacy of 'proper' randomisation in trials is as understandable as it was valuable.

With respect to the risks of alternate allocation, the primary one is surely the introduction of accidental bias as discussed earlier, or perhaps, with the same result, selection bias if assignment is not concealed. The potential for errors in variance estimation is, however, also important. The assumption of no bias must be combined with the assumption of an appropriate sampling frame for statistical inference if alternate allocation is to be appropriate.

Randomisation supplies, to use the terminology of Yates,³⁶ an 'objective' sampling frame, and the validity of significance tests and confidence intervals depends either on this or on an assumption that the data can be regarded as coming from a suitable sampling frame. As indicated in a personal communication to Iain Chalmers,³⁷ Armitage recognised that alternation can go wrong in this respect if successive responses are not statistically independent but he doubted 'whether the effect would be important'. In spite of this, one suspects that Armitage, and Mainland would have agreed that the safest approach is true randomisation.

The importance of an objective sampling frame is that it justifies the probability calculations used in significance tests. Independence of observations is critical and this is generally combined with distributional assumptions. Probability calculations may be numerically 'exact' as for Fisher's test for 2 × 2 tables based on binomial distribution assumptions, although these can be derived from first principles as well, or for *t*-tests or *F*-tests based on initial assumptions that the data are drawn from normal distributions. Mainland mentions such calculations throughout his book while also pointing out the situations when approximations are practically useful and accurate, for example, for chi-squared tests for contingency tables.

The additional advantage given through randomisation however is that significance testing can be performed based only on the randomisation distribution without the need for distributional assumptions. Fisher, it seems,

regarded this as less important practically as evidenced in his book *The Design of Experiments* where he writes (pp. 50–51 of the first edition),²⁷ in referring to a test of the difference in two normal means:

‘There has, however, in recent years, been a tendency for theoretical statisticians, not closely in touch with the requirements of experimental data, to stress the element of normality in the hypothesis tested, as if it were a serious limitation to the test provided. It is, indeed, demonstrable that, as a test of this hypothesis, the exactitude of “Student’s” *t*-test is absolute. It may, nevertheless, be legitimately asked whether we should obtain a materially different results were it possible to test the wider hypothesis which merely asserts that the two series are drawn from the same populations, without specifying that this is normally distributed’.

Frank Yates, Fisher’s close collaborator, writes consistently with this view, in a slightly different context, that questioning the validity of a random experiment ‘because the original material is not normally distributed’ must ‘be regarded rather as a debating point than a serious objection’ on page 441 of a 1939 article.³⁶

Rosenberger et al.,³⁸ noting current computing capabilities, nevertheless makes a strong case for the use of randomisation tests, even or even especially in more complex trials. Also, the use of a randomisation test might be seen as a simple example of ‘assumption lean inference’³⁹ which more generally argues for robustness of inference to allow for mis-specified models. In addition, minimally, as is also highlighted by Rosenberger, the examination of randomisation tests can be very helpful in understanding the nature of an experiment and the essential aspects of any analysis that is adopted, whether a formal randomisation test is used, or even practical. An earlier illustration of this is given by Nelder.⁴⁰ This is also consistent with the remark of Cox⁴¹ concerning clinical trials: ‘While the final analysis may not be based explicitly on the randomization distribution, it is necessary that there should be some broad correspondence with randomization theory’.

However, all these authors would also agree that some medical studies are, and may have to be, observational and do not involve randomisation of comparison groups. In these studies, the two problems of bias and error estimation cannot be removed by randomisation so other arguments must be made. These are explored in the next section.

Non-randomised studies and causal inference

On page 5 of *Elementary Medical Statistics*, Mainland says that there is ‘no absolute division between observation with experiment and observation without experiment’. Later in the same chapter on page 37, he deals specifically with the issue of ‘Causal Interpretation’.

While he describes this section as a ‘glance’ at the topic, it is noteworthy that the flavour of the discussion is totally consistent with the, now famous, criteria for causation in the presence of a demonstrable association that were given by Austin Bradford Hill in a 1965 article.⁴² This 1965 article builds on or is closely linked with prior work by Yerushalmy and Palmer,⁴³ the 1964 US Surgeon General’s report on Smoking and Health⁴⁴ and an earlier 1962 article by Hill.⁴⁵

Mainland’s discussion first acknowledges the issues raised due to the multiplicity of causes that may play a role in a disease process. He then specifically, referencing Greenwood⁴⁶ who wrote that any causal interpretation of an association must be credible biologically, acknowledges that this depends on the state of knowledge concerning the disease process, and, specifically, that the time relationship between putative cause and effect must be sensible. Summarising later, he describes this as showing ‘why’ a demonstrable association occurred.

It is certainly noteworthy that Mainland presents such a discussion well before it gained the prominence it did in the later part of the 1950s, through discussions of Doll and Hill’s work on the link between smoking and cancer. Parenthetically, it can also be noted that, in this discussion, Mainland alludes to the plausibility of a link between cigarettes and mortality from heart disease, a link which leads to more deaths than the lung cancer link although the relative risk is lower.

With respect to the statistical methods to be used in observational studies, Mainland writes, following his observation of no absolute division between observational and experimental data ‘One of the most important techniques in Fisher’s *Statistical Methods* was illustrated in the study of rainfall records; and the methods of sampling and analysis of observational data in public health are now being changed in accordance with the new methods’. Therefore, it is clear that Mainland accepted the validity of statistical methods, that is, that they have known error properties, in non-randomised studies, and this is separate to the issue of drawing causal inferences from such studies. It is not the validity of a significance test that is questionable, it is its meaning, that is, the broader inference, that can be drawn from them.

With respect to the sampling frame that underlies significance testing in observational studies, Fisher is sometimes characterised as a frequentist, a term that relates to the notion of the long-run behaviour of statistical procedures. Cox⁴⁷ however writes:

‘Fisher is often thought of as frequentist in his thinking, but this is rather misleading. He strongly emphasised that when probability was used to describe what underlay a set of data, he did not have in mind probability as a limiting frequency over a large number of repetitions. Rather, by probability Fisher meant a proportion in a hypothetical infinite population, the data being regarded as a random sample from

that hypothetical population. This in particular allowed the associated methods to be applied to situations, such as studies of literary authorship, in which direct replication of the data was inconceivable’.

Discussion

A remarkable career

In his biography of Mainland, Altman²⁵ wrote ‘Donald Mainland’s career was remarkable, with a unique move from anatomy to medical statistics and clinical trials’. Mainland is not unique among early medical statisticians in moving from medicine to medical statistics. Major Greenwood⁴⁸ and Raymond Pearl¹⁷ are other, earlier, examples. Indeed, while the pathway from medicine to medical statistics is not as evident currently, it is not unknown and probably still brings some particular advantages. However, there is a uniqueness in Mainland’s move from the field of anatomy where experimentation was a central feature of research. This background might be a factor that contributes in different ways to the continued relevance of Mainland’s book and was surely a factor in his recognition that Fisher’s work on experimental design and statistical inference should inform medical research.

The remarkableness of Mainland’s career is not solely linked to his appreciation of Fisher however. As evidenced in his treatment of randomisation and causal inference, Mainland was concerned with many aspects of the treatment of medical data, and this is even more evident when examining the full contents of his 1952 book which is done in the Appendix to this article⁵⁴⁻⁵⁶. For example, as Neuhauser et al.⁴⁹ and Matthews⁶ pointed out, factorial designs, which Fisher promoted, were identified as important by Mainland but little mentioned by other medical statisticians at that time. Furthermore, Mainland covers two topics that were little discussed at that time but which are currently major topics of interest in medical statistics. These are intercurrent events discussed in Chapter 4 and mixture models in Chapter 5. The brief remarks on these in the Appendix are included here as well.

Intercurrent events

There is a brief subsection in the section on experimental planning in Chapter 4 on what Mainland calls intercurrent events. This would appear to be the first use of this, now more commonly used, term, often in the context of discussions of intention-to-treat,^{50,51} to refer to the possibility of events taking place after treatment which may influence the patient’s outcome. The given examples of these events are treatment supplementation, change of treatment, accidents or diseases which may or may not be associated with the condition under study, the suspension of treatment for the patient’s business or domestic affairs

and loss of follow-up of a patient for a variety of possible causes including death. Mainland summarises what should be done as follows: ‘In deciding what should be done with data from any such patients the criterion must always be whether their inclusion or omission would introduce bias. Unless the appropriate decision is obvious, the best plan is to analyse all the data together, then to analyse the special cases and the main series separately’. The inclusion of such a section in 1952 seems particularly remarkable.

Mixture models

A short subsection of interest in Chapter 5, titled ‘Heterogeneous Samples’, makes the general point about being aware of known heterogeneity but focuses on the example of dental caries where patients with zero-caries may have a high or low frequency independent of the frequencies of non-zero categories of caries. This can be seen to be an early example of the possible value of mixture models for zero-heavy count data and other similar data.⁵²

Final remarks

As indicated previously, the writing of *Elementary Medical Statistics* was motivated by Mainland’s interest in the teaching of statistics to students of medicine. However, as discussed in Section 3, Mainland had reservations about its formal introduction into medical training. The debate on how medical statistics is best introduced continues in medical schools today although, due to Mainland and others, its importance is unquestioned. Reflection on Mainland’s writings can certainly contribute to this debate.

Finally, I should like to note the influence of Mainland on the late Tony Johnson⁵³ with whom I had the privilege of writing numerous articles on the history of medical statistics, some of which are referenced in this article. Tony entered the field of medical statistics from mathematics. A key book for him, when he needed a rapid introduction to the field, was *Elementary Medical Statistics* by Mainland, likely the second edition. It was therefore part of the foundation of Tony’s long career in medical statistics and, I am sure, the work of many others has also been influenced by the passion of Mainland in making statistical thinking a key component of medical research.

Declaration

Acknowledgements: I thank Iain Chalmers for his suggestion that Mainland’s 1952 book warranted a detailed examination and for his encouragement throughout the writing of this article. I also thank Robert Matthews, Daniel Farewell and Agnes Herzberg for helpful comments and discussions, and John Matthews for a careful and thoughtful review of the manuscript.

Competing interests: None declared.

Funding: None declared.

Ethical approval: Not applicable.

Guarantor: VF.

Contributorship: Sole authorship.

Provenance: Invited article from the James Lind Library.

Supplemental material: Supplemental material for this article is available online.

ORCID iD: Vern Farewell  <https://orcid.org/0000-0001-6704-5295>

Use of generative AI: No generative AI was used during the preparation of this manuscript.

References

- Chalmers I, Dukan E, Podolsky S and Davey-Smith G. The advent of fair treatment allocation schedules in clinical trials during the 19th and 20th centuries. *J Royal Soc Med* 2012; 105: 221–227.
- Medical Research Council (1948) Streptomycin treatment of pulmonary tuberculosis. *Br Med J* 2: 769–782. See <https://www.jameslindlibrary.org/medical-research-council-1948b/> (accessed 9 April 2026).
- Mainland D. *Elementary medical statistics*, 1st ed. Philadelphia, PA: WB Saunders Company, 1952.
- Matthews RAJ. The problematic history of randomised controlled trials, Part 1: presumption and confusion on the road to randomisation. *J R Soc Med* 2026; 119(1): 25–30. See <https://www.jameslindlibrary.org/articles/the-problematic-history-of-randomised-controlled-trials-part-1-presumption-and-confusion-on-the-road-to-randomisation/> (accessed 9 April 2026).
- Matthews RAJ. The problematic history of randomised controlled trials, Part 2: Hill's "pragmatic" view of randomisation and its origins. *J R Soc Med* 2026; 119(2): 55–58. See <https://www.jameslindlibrary.org/articles/the-problematic-history-of-randomised-controlled-trials-part-2-hills-pragmatic-view-of-randomisation-and-its-origins/> (accessed 9 April 2026).
- Matthews RAJ. The problematic history of randomised controlled trials, Part 3: Mainland, Hill and the future of RCTs. *J R Soc Med* 2026. DOI: 10.1177/01410768261419062. See <https://www.jameslindlibrary.org/articles/the-problematic-history-of-randomised-controlled-trials-part-3-mainland-hill-and-the-future-of-rcts/> (accessed 9 April 2026).
- Farewell V and Johnson T. Major Greenwood's early career and the first departments of medical statistics. *Stat Med* 2014; 33: 2161–2173. See <https://www.jameslindlibrary.org/farewell-v-johnson-t-2014/> (accessed 9 April 2026).
- Farewell VT and Johnson TL. Commentary: Dr John Brownlee MA, MD, DSc, DPH (Cantab), FRFPS, FSS, FRMetS (1868–1927), public health officer, geneticist, epidemiologist and medical statistician. *Int J Epidemiol* 2014; 42: 935–943.
- Farewell V, Johnson T and Armitage P. 'A Memorandum on the Present Position and Prospects of Medical Statistics and Epidemiology' by Major Greenwood. *Stat Med* 2006; 25: 2161–2177.
- Greenwood M. Accident proneness. *Biometrika* 1950; 37: 24–29.
- Armitage P. Bradford hill and the randomized controlled trial. *Pharm Med* 1991; 6: 23–37.
- Armitage P. Before and after Bradford hill: some trends in medical statistics. *J R Stat Soc Ser A* 1995; 158: 143–153.
- Armitage P. Fisher, Bradford Hill, and randomization. *Int J Epidemiol* 2003; 32: 925–928.
- Hill AB. *Principles of Medical Statistics*. London, UK: Lancet. See <https://www.jameslindlibrary.org/hill-ab-1937b/> (1937, accessed 9 April 2026).
- Farewell V and Johnson A. The first British textbook of medical statistics. *J Royal Soc Med* 2012; 105: 446–448. See <https://www.jameslindlibrary.org/articles/the-first-british-textbook-of-medical-statistics/> (accessed 9 April 2026).
- Farewell V and Johnson A. The origins of Austin Bradford Hill's classic textbook of medical statistics. *J R Soc Med* 2012; 105: 483–489. See <https://www.jameslindlibrary.org/articles/the-origins-of-austin-bradford-hills-classic-textbook-of-medical-statistics/> (accessed 9 April 2026).
- Greenwood M. Obituary: Raymond Pearl. *J R Stat Soc* 1941; 104: 94–96.
- Pearl R. *Medical Biometry and Statistics*, 1st ed. Philadelphia: WB Saunders Company, 1923.
- Pearl R. *Medical Biometry and Statistics*, 3rd ed. Philadelphia: WB Saunders Company, 1940.
- Cochran WG. The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *Math Proc Cambridge Philos Soc* 1934; 30: 178–191.
- Cox DF. Snedecor, George Waddell. In P. Armitage, T. Colton ed., *Encyclopedia of Biostatistics*, 5009–11. Chichester, UK: Wiley, 2005.
- Gehan EA and Lemak NA. *Statistics in Medical Research: Developments in Clinical Trials*. New York: Plenum, 1994.
- Lilienfeld DE. Harold Fred Dorn and the First National Cancer Survey (1937–1939): the founding of modern cancer epidemiology. *Am J Public Health* 2008; 98(12): 2150–2158.
- Cornfield J. Harold Fred Dorn. *Am Stat* 1963; 17: 53.
- Altman D. Donald mainland: anatomist, educator, thinker, medical statistician, trialist, rheumatologist. *J R Soc Med* 2020; 113: 28–38.
- Fisher RA. *Statistical Methods for Research Workers*. London, UK: Oliver; Boyd, 1925.
- Fisher RA. *The Design of Experiments*. London, UK: Oliver; Boyd, 1935.
- Mainland D. *The Treatment of Clinical and Laboratory Data: An Introduction to Statistical Ideas and Methods for Medical and Dental Workers*. Edinburgh, UK: Oliver; Boyd. See <https://www.jameslindlibrary.org/mainland-d-1938/> (1938, accessed 9 April 2026).
- Mainland D. Statistical methods in medical research I. Qualitative Statistics (Enumeration Data). *Can J Res* 1948; 26: 1–166. See <https://www.jameslindlibrary.org/mainland-d-1948/> (accessed 9 April 2026).
- Mainland D Sutcliffe MI. Statistical methods in medical research. II. Sample sizes required in experiments involving all-or-none responses. *Can J Med Sci* 1983; 53: 406–416.
- Cox DR. *Planning of experiments*. New York: Wiley. See <https://www.jameslindlibrary.org/cox-dr-1958/> (1958, accessed 9 April 2026).
- Mainland D. Review: controlled clinical trials, Ed: A. B. Hill. *J Am Stat Assoc* 1961; 56: 758–759.
- Hill AB ed. *Controlled Clinical Trials*. Springfield, IL: Charles C Thomas, 1960.
- Armitage P. Obituary: Sir Austin Bradford Hill, 1897–1991. *J R Stat Soc Ser A* 1991; 154: 482–484.
- Silverman WA and Chalmers I. Casting and drawing lots: a time-honoured way of dealing with uncertainty and ensuring fairness. *Br Med J* 2001; 323: 1467–1468. See <https://www.jameslindlibrary.org/articles/casting-and-drawing-lots-a-time-honoured-way-of-dealing-with-uncertainty-and-for-ensuring-fairness/> (accessed 9 April 2026).
- Yates F. The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments. *Biometrika* 1939; 30: 440–466.

37. 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: 1156–1164.
38. Rosenberger WF, Uschner D and Wang Y. Randomization: the forgotten component of the randomized clinical trial. *Statist Med* 2019; 38: 1–12.
39. Vansteelandt S and Dukes O. Assumption-lean inference for generalised linear model parameters (with Discussion). *J R Stat Soc Ser B* 2022; 84: 657–739.
40. Nelder JA. The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc R Soc London A* 1964; 283: 147–162.
41. Cox DR. A remark on randomization in clinical trials. *Utilitas Mathematica* 1982; 21A: 245–252.
42. Hill AB. The environment and disease: association or causation. *Proc R Soc Med* 1965; 58(5): 295–300.
43. Yerushalmy J and Palmer CE. On the methodology of investigations of etiologic factors in chronic diseases. *J Chronic Dis* 1959; 10: 27–40.
44. United States Public Health Service. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, DC: Department of Health, Education; Welfare, 1964.
45. Hill AB. The statistician in medicine. *J Inst Actuar* 1962; 88(2): 178–191.
46. Greenwood M. Mr. Shaw on doctors. *Br Med J* 1944; 2: 570–571.
47. Cox DR. Some pioneers of modern statistical theory: A personal reflection. *Biometrika* 2016; 103: 747–759.
48. Farewell V and Johnson T. Major Greenwood (1980-1949): The biography. *Statistics in Medicine* 2016; 35: 5533–35. See <https://www.jameslindlibrary.org/farewell-v-johnson-t-2016b-2/> (accessed 9 April 2026).
49. Neuhauser D, Provost SM and Provost LP. It is time to reconsider factorial designs: how Bradford Hill and RA Fisher shaped the standard of clinical evidence. *Qual Manag Health Care* 2020; 29: 109–122.
50. Chalmers I, Matthews R, Glasziou P, et al. Trial analysis by treatment allocated or by treatment received? Origins of the ‘Intention-to-Treat Principle’ to Reduce Allocation Bias: Part 1. *J R Soc Med* 2023; 116: 343–350. See <https://www.jameslindlibrary.org/articles/analysis-of-clinical-trial-by-treatment-allocated-or-by-treatment-received-applying-the-intention-to-treat-principle/> (accessed 9 April 2026).
51. Chalmers I, Matthews R, Glasziou P, Boutron I and Armitage P. Trial analysis by treatment allocated or by treatment received? Origins of the ‘intention-to-treat principle’ to reduce allocation bias: part 2. *J R Soc Med* 2023b; 116: 386–394. See <https://www.jameslindlibrary.org/articles/analysis-of-clinical-trial-by-treatment-allocated-or-by-treatment-received-applying-the-intention-to-treat-principle/> (accessed 9 April 2026).
52. Farewell VT, Long DL, Tom BDM, Yiu S and Su L. Two-part and related regression models for longitudinal data. *Annu Rev Stat Appl* 2017; 4: 283–315.
53. Farewell V. Obituary: Anthony Leonard Johnson (1943–2022). *Statist Med* 2023; 42: 597–599.
54. Morgan KL and Rubin DB. Rerandomization to improve covariate balance in experiments. *Ann Stat* 2012; 40: 1263–1282.
55. Snedecor GW. *Statistical Methods Applied to Experiments in Agriculture and Biology*. Ames, IA: Iowa State College Press, 1946.
56. Yates F. Contribution to discussion of ‘The validity of comparative experiments’ by FJ Anscombe. *J R Stat Soc Ser A* 1948; 111: 204–205.