The origins of Austin Bradford Hill's classic textbook of medical statistics
Vern Farewell and Anthony Johnson

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
The Royal Society of Medicine

Additional services and information for Journal of the Royal Society of Medicine can be found at:

Email Alerts: http://jrs.sagepub.com/cgi/alerts
Subscriptions: http://jrs.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Nov 1, 2012

What is This?
The origins of Austin Bradford Hill’s classic textbook of medical statistics

Vern Farewell1 • Anthony Johnson1,2

1MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge CB2 0SR, UK
2MRC Clinical Trials Unit, Aviation House, 125 Kingsway, London, WC2B 6NH, UK
Email: vern.farewell@mrc-bsu.cam.ac.uk; tony.johnson@mrc-bsu.cam.ac.uk

Between January and April 1937 The Lancet published a series of seventeen articles on medical statistics by Austin Bradford Hill. The series was suggested by Dr MH (Pamela) Kettle, an assistant editor at The Lancet, and Hill was paid £3.3.0 (three pounds, three shillings, and no pence – roughly equivalent to £135 today) for each ‘installment’. The seventeen articles formed the chapters of Principles of Medical Statistics, which was also published in 1937. Over the next fifty-five years, the book was to run through twelve editions, and become renowned worldwide amongst physicians, epidemiologists, and medical statisticians.

Hill had a very distinguished ancestry. ‘From James (b. 1724) and Sarah Hill (∼1733–1801) descended one of the most intellectual families that has ever arisen in England’ (Gun, quoted by Hill). The family included Thomas Wright Hill (Hill’s great-great grandfather, 1763–1851), an educational reformer who devised a system of shorthand, and wrote on the hexadecimal system of counting and proportional representation in parliamentary elections; Sir Rowland Hill (1795–1879), who introduced ‘the penny post’; Matthew Davenport Hill (1792–1872), a criminal-law reformer; George Birkbeck Hill (Hill’s grandfather, 1835–1903), writer and critic renowned for his edition of Boswell’s Life of Johnson; Rosamund Davenport Hill (1825–1902), a prominent member of the old London School Board; Sir Maurice Hill (1862–1934), a High Court judge; and Sir Leslie Scott (1869–1950), a politician and judge who was briefly the Solicitor General. (He was the eldest son of Sir John Scott (1841–1904), another distinguished judge, whose sister, Annie (1830–1902), was the wife of George Birkbeck Hill; George’s sister, Laura, was the second wife of Edward Scott, Sir John’s father). Hill’s father, Sir Leonard Erskine Hill (1866–1952), became professor of physiology at the London Hospital (where he recruited Major Greenwood, widely regarded as the first medical statistician in Britain, in 1905), and, from 1914, head of the Department of Applied Physiology of the newly-formed Medical Research Committee (fore-runner of the Medical Research Council).

Unfortunately there is no biography of Austin Bradford Hill, although brief descriptions of aspects of his life, career, and contributions to statistical and medical research are available, for example, Doll, Farewell and Johnson, Gehan and Lemak, Keating, and Statistics in Medicine.

In 1927 Hill joined Major Greenwood’s department at the London School of Hygiene and Tropical Medicine (LSHTM), having obtained an honours degree in economics at University College London in 1922, and then a grant from the Medical Research Council to examine the high mortality in young adults in rural areas of England. While carrying out this study he attended Karl Pearson’s course on statistics at London University. Hill’s first four papers, for all of which he was the sole author, were published in 1925. The following year he was awarded a PhD by the University of London with a thesis entitled ‘A physiological and economic study of the diets of workers in rural areas as compared with those of workers resident in urban areas,’ which consisted of these four papers (we have not been able to trace the identity of Hill’s examiners). In 1930 Hill obtained a DSc from the University of London, with a thesis entitled ‘An investigation of sickness in various industrial occupations.’ This comprised eight
single-authored) papers published between 1925 and 1929, including the four submitted for his PhD degree. His external DSc examiner was WP Elderton, author of *Frequency curves and correlation* (1906) and *Primer of statistics* (1906). Major Greenwood was the internal examiner.

By 1936 Hill had published thirty-nine book reviews, eight research reports, and sixteen papers, including nine in the *British Medical Journal* or *The Lancet*, and four in the *Journal of the Royal Statistical Society*. Many of the reviews were of books about population, poverty, industrial working conditions, migration, mortality, and the social conditions in London, subjects of obvious central interest at the LSHTM. However Hill also reviewed Hartshorne and May’s *Studies in Deceit*, Burgess’s *Introduction to the Mathematics of Statistics*, Fisher’s *The Genetical Theory of Natural Selection*, and the second edition of Pearl’s *Introduction to Medical Biometry and Statistics*, thus demonstrating an interest extending into statistics, genetics, and beyond. Most of Hill’s papers also reflected a focus on mortality, longevity, social conditions, and industrial working conditions (including those of nurses), reflecting the rapid development of analyses of vital statistics which had occurred over the previous one hundred years. But his interests extended beyond mortality to cricket, experimental epidemiology, and medical practice.

### The Lancet articles

In 1936, Pamela Kettle, an assistant editor at *The Lancet*, invited Hill to write a series of articles for the Journal on medical statistics. We have no knowledge or record of why Hill was chosen to write the articles. By 1936 the number, scope, quality and clarity of his publications were readily apparent. The first article in the series was heralded by a *Lancet* editorial, entitled *Mathematics and Medicine*, in which the writer (possibly Greenwood) noted the author’s ‘wide experience of the problems with which medical men have to deal, and of the difficulties which post-graduate medical students encounter in grasping statistical principles.’ Editorial staff at *The Lancet* would have been aware of the need for better understanding of statistical methods through their own manuscript review process, and it is likely that staff in Greenwood’s department at LSHTM would have played a prominent role in this. In such circumstances, and with the second edition of Woods’ and Russell’s *An Introduction to Medical Statistics* having been published in 1936, Greenwood’s department would be an obvious place for *The Lancet* to seek an author. However, whether the invitation went initially to Greenwood or to Hill, or indeed to both, remains a matter of conjecture. By 1937 Greenwood himself had published five books, including an introduction to epidemiology, 21 reports, and over 100 papers.

Where did Hill develop the ideas that were to lay the foundations for this series of articles and for the famous book that was based on them? Did his ideas develop from his own research studies, his teaching, or perhaps through discussions with, or the publications of, his colleagues?

The influence of the first foundation – his research studies – was made clear in the Preface to the first edition of the book, where Hill remarks:

> The worker in medical problems, in the field of clinical as well as preventive medicine, must himself know something of statistical technique, both in experimental arrangements and in the interpretation of figures. To enable him to acquire some knowledge of this technique I have tried to set down as simply as possible the statistical methods that experience has shown me to be most helpful in the problems with which medical workers are concerned. I have used examples taken from medical inquiries in the attempt to make clear these methods of analysis, and have sought to show by illustration where and why workers make mistakes in their interpretation of figures.

What about the second possible foundation - Hill’s teaching at the LSHTM? In the decade from 1927, Hill would have taught statistics to medical students at LSHTM, although we cannot be sure exactly what these courses covered. He does not refer to them in the Preface to his book, although Sir Richard Doll did link them to the book in his entry on Hill in the Encyclopedia of Biostatistics. It seems certain, however, that Hill must have drawn somewhat on his teaching experience.
What of the third possible foundation - his collegiate links within LSHTM? We have discussed this in detail in another article. In summary, Hill was part of the more general statistics community as evidenced by his involvement with the Royal Statistical Society. In addition, he would have been directly influenced by Greenwood, who headed the statistics department at the LSHTM, and other staff colleagues, including, for example, Joseph Oscar Irwin, who is regarded as the leading theoretician among UK medical statisticians in the 1930s. Finally, when Hill was writing his articles for The Lancet, the text of An Introduction to Medical Statistics, by his colleagues Hilda M Woods and William T Russell, would have been in use at LSHTM.

Reactions to the series of papers in The Lancet were very complimentary and Hill received several requests for copies, all of which he declined initially because ‘the journal does not provide me with any reprints unless I purchase them.’ However, The Lancet did respond to at least one of these requests (that from Dr Claus Jensen, Director of the Department of Biological Standards at the State Serum Institute in Copenhagen) by sending complimentary copies of the journal as each article appeared. Once publication as a book was agreed, Hill changed his reason to ‘reprints are not available as the articles will be published as a book.’ Indeed some correspondents suggested publication as a book at an early stage and were enthusiastic about it.

R.A. Fisher (letter 9 April 1937) was ‘very glad to hear about your book. It, and probably others like it, are certainly much needed.’ Raymond Pearl, an American pioneer of statistical thinking in medicine, in a letter to Greenwood (12th April 1937), was effusive: ‘Please tell Hill that I think his series of articles on the Principles of Medical Statistics is superb. It is by long odds the best thing of its kind that has ever been done. Surely he plans to publish the series as a book. He must by all means do so.’

However Fisher’s letter also alarmed Hill somewhat by asking ‘I do not know if you know Donald Mainland, who has, I understand, a work on tests of significance in medicine now in the hands of the publishers.’ Hill replied (12th April 1937) that he did not know Donald Mainland and asked ‘is he the man who had an article in the BMJ some time back?’ He repeated the question to Pamela Kettle at The Lancet and continued ‘whoever he may be his appearance is rather annoying, I leave it to you to use as an argument for speed or to know nothing about.’

Mainland did indeed publish a paper in the BMJ on ‘Problems of chance in clinical work’ and acknowledged Fisher’s help in checking it. Mainland states in the summary that:

‘in medical work there is serious neglect of highly important problems of chance and of variation between samples, exemplified in a recent estimate of chances of cure in facial paralysis where odds are wrongly estimated, and too great reliance is placed on samples. What is required in clinical work is not elaborate mathematical tests, but an understanding of the meaning of chance, and adequate precautions that the samples, however small, are unbiased.’

Hill would undoubtedly have agreed.

Mainland’s article focuses from the start on the work on Bell’s palsy by Tumarkin (BMJ 21 March 1936, pages 580–581) though it was used ‘simply as an example of ideas and methods applicable to many problems.’ The criticism provoked an irritable response from the author, Tumarkin (BMJ letter 12th December 1936, pages 1229–1230), to which Mainland responded (BMJ 23 January 1937, pages 192–193). To his credit, Tumarkin was one of the first to request reprints of Hill’s articles (reported by Hill in a letter to Sir Squire Sprigg dated 26th January 1937).

Mainland did publish several texts on medical statistics starting in 1938 with The Treatment of Clinical and Laboratory Data: an Introduction to Statistical Ideas and Methods for Medical and Dental Workers, the text most likely to have alarmed Hill.

The Book

What was the genesis of the book itself? Hill’s correspondence concerning the book is in the collection of letters held in the MRC Biostatistics Unit in Cambridge. From a letter dated 26 January 1937, we learn that, at some stage in 1936, Hill had made an ‘only half-serious request’ to consider publication of the articles as a book. This suggestion was encouraged by
Sir Squire Sprigge, Editor of *The Lancet* (letter dated 16th December 1936), who believed that *The Lancet’s* publishers ‘would probably take the book at once’, and that he was prepared to tell them that ‘the subject is one that particularly requires some fundamental elucidation.’ However, two months later reservations were being expressed, with FGH Holt (Secretary at *The Lancet* Ltd., letter dated 25th February 1937) reporting that *The Lancet’s* directors (R Percy Hodder-Williams, Ralph Hodder-Williams and Garfield H Williams) were ‘apprehensive of the production and success of a small book at a low selling price’. Hill’s response was immediate (26th February 1937):

I should like to know as early as possible whether the Directors are willing to publish my articles in book form. I certainly understood from my correspondence with the Editor that that was the intention and I took a very great deal of trouble over the articles with that end in view. (A number of people have expressed to me the hope that the articles will be put into a book). If the Directors however feel that its success is too speculative I should wish to approach another publisher. [It is not known if this was the exact content of the letter that was sent; the above quotation is from a draft written at the foot of the letter from FGH Holt, who refers to Hill’s letter of 26th February in his subsequent letter of 5th April 1937].

After another two months *The Lancet’s* Directors relented, although with reservations and conditions (letter from FGH Holt dated 5th April 1937), as the Directors considered it ‘improbable that financial success would be achieved, and inferred that a period of time would elapse before even the cost of production could be recovered, apart from the substantial expense of publicity’; the conditions amounted to an estimated number of between 500 and 1,000 copies being sold without royalty payments to Hill.

Hill modified these terms (letter dated 9th April 1937), preferring a fixed figure to one based upon an ‘approximate estimate of the cost of production.’ He suggested foregoing royalties on the first 500 copies, then 10% of the published price on (ordinary) sales from 500 to 1,500 copies, of 12½% from 1,500 to 2,000 copies and 15% on all subsequent sales. *The Lancet* agreed to these terms, despite the cost coming out ‘more than [Holt] expected’ (letter from FGH Holt dated 26th April 1937). Perhaps this was just as well since the publisher, Oliver and Boyd, were expressing an interest (letter 13th April 1937).

In retrospect, the cautions of *The Lancet* Directors were somewhat misguided. Raymond Pearl was far more prescient and wrote to Hill (letter dated 5th May 1937) ‘in my opinion they *The Lancet* people are quite needlessly worried about the extent of its sale. Your presentation is excellent and I am sure the book will have a steady sale for a long time.’

The first edition of Hill’s book was published on 17th June 1937, less than two months after publication of the final Lancet article, although, as shown in Table 1, the order of presentation was changed.

Originally the last two articles were intended only for the book. Hill commented (letter to Sir Squire Sprigge, 26 January 1937):

In dealing with standard deviations and correlations I deliberately directed my discussion to the use of these measures and their meanings. I did not think it right, or wise, to fill up your columns with demonstrations of how they are in fact calculated. For this I referred readers to other text books, e.g. Woods and Russell. For the purpose of a book, however, I think it might be as well to have it more self-contained. If you agree I could add an appendix to the book (or chapters at appropriate points) showing as simply as possible how such things can be most conveniently calculated from actual data.

Pamela Kettle thought otherwise ‘because (1) some of our less affluent readers may be cutting out these articles to make a book of their own, and (2) at future times, when students or doctors are looking at the bound volumes of *The Lancet* they will want to make full use of the statistical articles without having to turn elsewhere.’

The book is dedicated to F.M.H (Hill’s wife, Florence Maud Hill), and has a Foreword by the Editor of *The Lancet*, and a Preface by Hill. It also included a table of $\chi^2$, reproduced from RA Fisher’s *Statistical Methods for Research Workers*, some definitions (common statistical terms, standard errors, correlation, rates, life table symbols), and an index (carefully compiled by Hill).
The Preface has acknowledgements to Major Greenwood, WWC Topley, JO Irwin, WT Russell (see reference 5) as well as to MH Kettle. Among others, the text has references to two statistical books, Pearl’s Medical Biometry and Statistics (2nd edition 1930), and the second edition of Woods’ and Russell’s An Introduction to Medical Statistics.10

The first four editions of Hill’s book have ‘Postgraduate series v. 3’ printed on the spine. We have not been able to trace any volumes after Hill’s in the series. The first volume in the series (published in 1931) was Eric Graham Howe’s Motives and Mechanisms of the Mind: an Introduction to Psychopathology and Applied Psychology, the title page of which had an interesting quotation from Ecclesiasticus (38,15): ‘He that sinneth before his Maker, let him fall into the hands of the physician.’ The book comprised twelve lectures given during the autumn of 1930 under the auspices of the Tavistock Square Clinic, which had been published as a series in The Lancet. The second volume – The Preventive Aspects of Medicine – which was published in 1934, contained a series of lectures delivered at King’s College Hospital Medical School during 1933–34, which had also been originally published in The Lancet. The lectures had been delivered by distinguished authorities including Sir George Newman, Sir Leonard Hill, Major Greenwood (on Epidemiology and its lessons), Millais Culpin, and WWC Topley (all except the first of these appear in the glossary in our article5 on Greenwood’s Memorandum on the Present Position and Prospects of Medical Statistics and Epidemiology.

Only two years after the publication of the first edition of Hill’s book, a second edition was launched. The principal change from the first edition was the addition of chapter 18 on the ‘Calculation of Standardised Death-Rates.’ Thereafter new editions appeared every six years on average, some with additional chapters, rearrangement, and extensive editing, as summarised in Table 2.

The last edition to retain the original name was the 9th, published in 1971. In his preface, Hill reiterated, from his preface to the 6th edition, that contemporary examples are not necessary for illustrative examples ‘but possibly it may appear to some readers to be important, and an air of modernity can at least do no harm.’ However, he remained keen not to change what had been successful and therefore ‘decided to leave alone much that seemed well.’ Despite this, the 9th edition is the largest edition of Hill’s book, with 24 chapters (30% up on the first edition), and 390 pages (more than double the 171 pages of the

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Comparison of The Lancet papers and the 1st edition of Principles of Medical Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Lancet articles</td>
<td>Book chapter</td>
</tr>
<tr>
<td>Number</td>
<td>Date (1937) and pages</td>
</tr>
<tr>
<td>1</td>
<td>2 Jan; 41–43</td>
</tr>
<tr>
<td>2</td>
<td>9 Jan; 99–101</td>
</tr>
<tr>
<td>3</td>
<td>16 Jan; 161–163</td>
</tr>
<tr>
<td>4</td>
<td>23 Jan; 219–221</td>
</tr>
<tr>
<td>5</td>
<td>30 Jan; 281–284</td>
</tr>
<tr>
<td>6</td>
<td>6 Feb; 337–340</td>
</tr>
<tr>
<td>7</td>
<td>13 Feb; 402–405</td>
</tr>
<tr>
<td>8</td>
<td>20 Feb; 459–461</td>
</tr>
<tr>
<td>9</td>
<td>27 Feb; 527–529</td>
</tr>
<tr>
<td>10</td>
<td>6 Mar; 583–586</td>
</tr>
<tr>
<td>11</td>
<td>13 Mar; 646–648</td>
</tr>
<tr>
<td>12</td>
<td>20 Mar; 706–708</td>
</tr>
<tr>
<td>13</td>
<td>27 Mar; 771–773</td>
</tr>
<tr>
<td>14</td>
<td>3 Apr; 825–827</td>
</tr>
<tr>
<td>15</td>
<td>10 Apr; 883–885</td>
</tr>
<tr>
<td>16</td>
<td>17 Apr; 941–943</td>
</tr>
<tr>
<td>17</td>
<td>24 Apr; 1001–1003</td>
</tr>
</tbody>
</table>
Some concluding reflections

Hill’s book was innovative and useful to a wide range of readers, and was translated into Spanish (1st ed. 1950; 2nd ed. 1958; 3rd ed. 1965), Russian (1958), Korean (1960), and Polish (1962). We have examined its broader historical background elsewhere.12 Briefly, early texts with medical statistics in their title appeared in the 1820s and we identified 57 textbooks on vital and medical statistics published by 1930. These included the US publication by Raymond Pearl18 titled Introduction to Medical Biometry and Statistics. We concluded that ‘Hill’s book stands out as making a worldwide contribution to the understanding and teaching of medical statistics over the last 70 years. It remains the best known text on the subject.’

A comprehensive study to track all the changes throughout the twelve editions, might offer some revelations on the development of medical statistics more generally after 1937 as suggested by Horton.19 We note that Chalmers20 has done this already with his critical examination of Hill’s description and discussion of treatment allocation in clinical trials. It would be of particular interest to assess the influence of Hill’s book, if any, on later statistical texts.

For example, Mainland’s The Treatment of Clinical and Laboratory Data: an Introduction to Statistical Ideas and Methods for Medical and Dental Workers17 was published the year after the first edition of Hill’s book, but does not mention either Hill’s Lancet articles or his book. Other early examples include Statistical Methods for Medical and Biological Students by Gunnar Dahlberg (1940), Statistics for Medical and other Biological Students by Bernstein and Weatherall (1952), and Elements of Medical Statistics by JV Smart (1963).

The aim of texts published after the first edition of Hill’s book, their content, their perception, and the need for them, together with an appraisal of why so few went beyond a first edition, might indeed demonstrate that the influence of Hill’s book went far beyond its immediate content. The publication history of Hill’s book shown in Table 2 is sufficient to indicate the special place this work has in the history of medical statistics.
and the training of doctors, epidemiologists and statisticians.

References

2 Hill ID. A commemorative issue devoted to Austin Bradford Hill. Statistics in Medicine 1982;1:4
15 Pearl R. A statistical discussion of the relative efficacy of different methods of treating pneumonia. Archives of Internal Medicine 1919;24:398–403
17 Mainland D. The treatment of clinical and laboratory data: an introduction to statistical ideas and methods for medical and dental workers. Edinburgh: Oliver and Boyd, 1938