History of biostatistics

J. Rosser Matthews
University of Maryland, College Park, MD, USA

Correspondence to:
J. Rosser Matthews
Professional Writing Program
English Department
1220 Tawes Hall
University of Maryland
College Park, MD, USA
+1 (301) 405-3762
jrmatt3@umd.edu

Abstract
The history of biostatistics could be viewed as an ongoing dialectic between continuity and change. Although statistical methods are used in current clinical studies, there is still ambivalence towards its application when medical practitioners treat individual patients. This article illustrates this dialectic by highlighting selected historical episodes and methodological innovations – such as debates about inoculation and blood-letting, as well as how randomisation was introduced into clinical trial design. These historical episodes are a catalyst to consider assistance of non-practitioners of medicine such as statisticians and medical writers.

Methodologically, clinical trials and epidemiological studies are united by a population-based focus; they privilege the *group* (i.e., population) over the clinically unique individual. Over time, this population-based thinking has remained constant; however, the specific statistical techniques to measure and assess group characteristics have evolved. Consequently, the history of biostatistics could be viewed as an ongoing dialectic between continuity and change. The continuity derives from focusing on the group rather than the clinically distinct individual. The change derives from developments in statistical theory that have led to more sophisticated analyses. In this article, I will illustrate this dialectic by discussing examples from antiquity to the emergence of the clinical trial in the mid-20th century.

Ancient sources: Hippocratic writings and the bible

Although the Hippocrates writers (active in the 5th century BCE) did not employ statistical methods, one treatise does stand out as a pioneering example of an environmental epidemiological study– the treatise *On Airs, Waters, and Places* (c. 400 BCE).[1] Relying on a view of disease as based on an imbalance in bodily fluids– known as *humours* – the work emphasised how climatic changes throughout the seasons of the year contributed to the spread of different types of diseases.[1] While basically qualitative, the work is historically significant because it looked beyond the individual to suggest a role for larger geographic and environmental factors. Furthermore, it relied on naturalistic explanations rather than invoking various deities to account for illness and therefore anticipated a modern scientific outlook.

Another ancient forerunner of contemporary clinical trials is discussed in the Bible's *Book of Daniel*. King Nebuchadnezzar of Babylon wanted all of his subjects to eat a diet of only meat and wine. However, Daniel and some of the other Jewish children wanted to eat a diet of legumes and water. The King permitted them this diet for 10

days – after which it was determined that they were indeed healthier. Consequently, they were allowed to continue on this diet.[2] Although not having the "apparatus" of a modern clinical trial (e.g., statistical tests to determine p-values, confidence intervals etc.), it does illustrate the use of a comparison to test the efficacy of a dietary intervention.

Eighteenth century developments

In the 18th century, one prominent example of using statistical methods to resolve therapeutic debates centred on the practice of smallpox inoculation. This involved inserting actual smallpox pustules under an individual's skin in the hope of creating a mild (i.e., non-disfiguring) case of the disease that would induce later immunity. Since this actually put patients at risk of contracting a potentially fatal form of the disease, this became the subject of much controversy.

Some argued against this procedure based on the Hippocratic injunction "first, do not harm." However, many writers justified the procedure based on arguments that today would be called "risk-benefit analysis." For example, the London physician John Arburthnot (1665-1735) published an anonymous pamphlet in 1722, in which he examined the *London Bills of Mortality* from earlier years and estimated that the chance of dying from naturally-occurring smallpox was 1:10. He then asserted (without evidence) that the chance of dying from inoculation-induced smallpox was 1:100. This ten-fold reduction made him conclude that inoculation made sense: "A Practice which brings the Mortality of the Small Pox from one in ten to one in a hundred, if it obtain'd universally would save the City of London at least 1,500 People Yearly; and the same Odds wou'd be a sufficient prudential Motive to any private Person to proceed upon."[3] In 1760, a more mathematically sophisticated version of this type of analysis took place in a debate between the Swiss mathematician Daniel Bernoulli (1700-1782) and the French mathematician Jean d'Alembert (1717-1783). Bernoulli drew on probability mathematics to contrast life expectancies for inoculated and non-inoculated individuals; also, he calculated the benefits of inoculation broken down by age. D'Alembert challenged Bernoulli's assumptions and said that Bernoulli's model had not accurately captured the psychology of human decision making – would an individual accept the risk of death now (from inoculation) for an expected "pay-off" of additional years of life when one was old and feeble?[3]

While the debates about inoculation relied on mortality statistics, the individual that is more often credited with designing a controlled clinical trial (i.e., intentionally dividing the participants into two or more comparable groups to test hypotheses) is James Lind (1716-1794). In 1757, Lind (a ship's surgeon) had to deal with an outbreak of scurvy. He selected 12 of the sailors and divided them into six groups of twos. All were given the same diet – except for a key different ingredient for each of the distinct six groups. For the two sailors who received oranges and limes as supplement, there was one complete and one near recovery; none of the other five groups improved as much. Despite some obvious structural similarities to the Biblical account, Lind is today regarded as the (modern) "father" of the controlled clinical trial.[2]

Nineteenth century developments

In several areas of 19th century scientific endeavour, statistical reasoning was introduced – and the field of medicine was no exception. In the 1830s, one of the most prominent advocates for applying the "numerical method" to medicine was the French clinician Pierre-Charles-Alexandre Louis (1787-1872) (Figure 1). By collecting data on patients admitted to hospitals, Louis argued that the practice of bloodletting was actually doing more harm than good. In his 1835 treatise *Recherches sur les effets de la saignée*, Louis pointed out that 18 patients died out of the 47 who had been bled (approximately 3:7) whereas only nine died out of the 36 patients not bled – producing a lower mortality rate of approximately 1:4.[4]

Louis justified his approach by claiming that the difference between numbers and words (such as "more or less" and "rarely or frequently") is "the difference of truth and error; of a thing clear and truly scientific on the one hand, and of something vague and worthless on the other." Furthermore, Louis prophesied that, with the widespread introduction of numerical reasoning, "we shall hear no more of medical tact, of a kind of divining power of physicians."[4] In language that foreshadows contemporary discussions of "evidence-based medicine," Louis was basically saying that the key to transform medicine into a science was to rely on population-based thinking rather than individual expertise.

Some of Louis' contemporaries criticised his approach for failing to acknowledge that the physician had to treat the individual as a patient rather than a statistical construct. For instance, the physician Benigno Risueño d'Amador (1802-1849) used an analogy to maritime insurance. Although past experience might tell you that 100 vessels would perish for each 1,000 that embarked, these population-based regularities could not tell you which specific ships

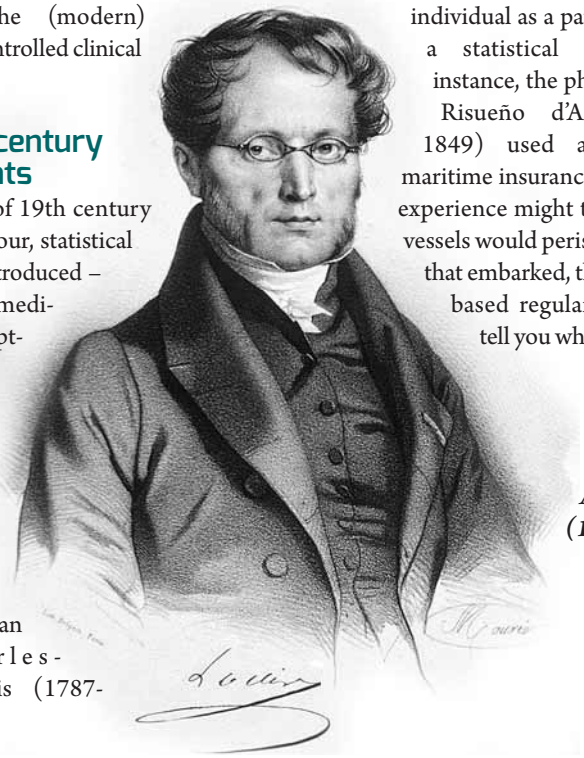

Figure 1. Pierre-Charles-Alexandre Louis (1787-1872) was a pioneer of the "numerical method" in medicine.

would be destroyed. Analogously, Risueño d'Amador argued that the calculus of the mathematicians "cannot be used to forecast a determined event, but only to establish the probability of a certain numerical proportion between two classes of possible events. But it is precisely this fact which makes it completely useless in medicine."[5] Drawing a different analogy, the physician François Double (1776-1842) claimed that relying on the numerical method would reduce the physician to "a shoemaker who after having measured the feet of a thousand persisted in fitting everyone on the basis of the imaginary model."[6]

These types of criticisms discussed statistical reasoning in the context of medical ethics: should the physician be concerned primarily with advancing scientific knowledge (through collecting empirical data), or with treating the individual in need of medical care? At the same time, however, a more mathematically sophisticated critique of Louis' work was developed by the French physician Jules Gavarret (1809-1890) (Figure 2), who had been trained as an engineer before becoming a physician and therefore understood probability mathematics. Gavarret published a treatise in 1840 entitled *Principes généraux de statistique médicale* in which he pointed out that Louis's averages could vary between what he called "limits of oscillation" if multiple samples were taken from the same population. For instance, Louis had observed 140 cases of typhoid fever with 52 deaths and 88 recoveries, or a mortality of 37%. Relying on probabilistic considerations, however, Gavarret posited that the results could vary by 11.55%, or between 26% and 49% in every 140 cases observed.[7] In modern day parlance, Gavarret was reporting the "confidence interval" associated with Louis' result.

To modern eyes, Gavarret seems remarkably prescient; however, there was no receptive audience for this marrying of statistics to probability mathematics in mid-19th century medicine. While his treatise was commented on throughout the 19th century (with varying degrees of mathematical sophistication), no "school" of followers committed to Gavarret's specific mathematical approach emerged. As a result, the meaning of statistical evidence remained contentious throughout the 19th century. For example, the famous surgeon, Joseph Lister (1827-1912), argued for his particular method of antiseptic surgery based on statistical studies; however, his critics had alternative theories of how to make surgery safer, citing other statistical studies that claimed to establish the superiority of their alternative theoretical approaches.[8]

The creation of the modern clinical trial

The move to standardise and "mathematise" statistics came with the creation, at University College London, of the Biometric School in 1893 and the Biometric Laboratory a decade later.[9] Heading the School and Laboratory was the pioneering statistician Karl Pearson (1857-1936) (Figure 3) who developed many modern statistical techniques to study biological variation – such as curve-fitting and goodness-of-fit tests, as well as methods for measuring correlation.[9] While Pearson's interest in developing statistics derived from a desire to make explicit the statistical implications of Darwin's theory of natural selection, he also advocated the extension of these methods into medicine. To that end, he often contributed to the *British Medical Journal*, the *Lancet*, and *The Royal Society of Medicine* as attempts to "educate" the medical profession on the proper methods of statistical reasoning.[10]

One physician who would actively embrace Pearson's recommendations was Major Greenwood (1880-1949). Greenwood studied under Pearson in 1904-1905 at the same time that he received his licence to practice medicine. At the beginning of 1910, Greenwood would be awarded a full-time position as a medical statistician at the Lister Institute of Preventive Medicine. Like Pearson, Greenwood would proselytise for statistical methods by debating with physicians. One of his most noteworthy encounters involved an exchange in the *Lancet* in 1912-1913 with the bacteriologist Sir Almroth Wright (1861-1947) over Wright's use of vaccines to combat pneumococcal infection among South African mine workers. Centring on the issue of the accuracy of a blood test, the debate evolved into a more generalised discussion over which forms of scientific evidence were more credible.[11]

By forging a career in academic science, Greenwood would help lay the foundations for a mathematically-informed understanding of epidemiology. In 1927, he would become the first professor of epidemiology and vital statistics at the London School of Hygiene and Tropical Medicine (LSHTM);[12] in 1928, he would be elected to the Royal Society. Also, Greenwood would train many students, of which the most prominent would be Sir Austin Bradford Hill (1897-1991).

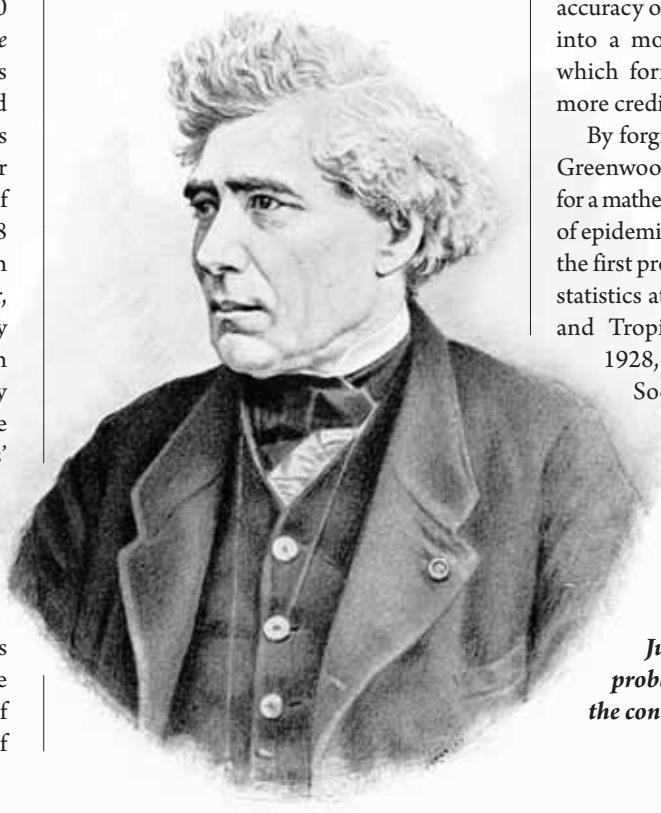

Figure 2. Jules Gavarret (1809-1890) used probability mathematics by applying the concept of the confidence interval to medical statistics.

Figure 3. Karl Pearson (1857-1936) developed curve-fitting methods and measures of correlation.

Bradford Hill was the third son of the physiologist Sir Leonard Hill (1866-1952) and had planned on following his father's medical profession. However, he contracted tuberculosis during World War I, and eventually earned an economics degree by correspondence. Hill gravitated towards statistics attending Pearson's lectures. In 1933, he would be appointed to a Readership at the LSHTM; upon Greenwood's retirement in 1945, Hill would succeed him as the head of the Statistics and Epidemiology Unit.[9] Like Pearson and Greenwood, Hill sought to educate the medical profession on the proper use of statistics. In 1937, he wrote a series of articles explaining statistical methods for the *Lancet*; subsequently, they would be republished as *Principles of Medical Statistics* and go through multiple editions and translations. In 1946, Hill would design a famous clinical trial to test the efficacy of streptomycin in treating tuberculosis – a methodologically noteworthy trial because it used a series of random sampling numbers to assign patients to the control (bed rest) or experimental (streptomycin) group. This trial is often characterised as the first clinical trial to use a randomisation scheme effectively. In 1965, Hill would articulate what have come to be known as the "Bradford Hill Criteria." These criteria can be used to determine whether an empirically observed association (e.g., between cigarette smoking and cancer) might be suggestive of an underlying causal relationship.

Today, the clinical trial is held as the gold standard for certain knowledge, and statistically-based epidemiological studies are widely reported in the news. However, as this brief historical sketch has illustrated, the current ascendency of these population-based thinking masks a larger ambivalence towards statistical methods within the medical profession. Even as statistical methods have been used to justify notable therapeutic breakthroughs, the population-based thinking on which they are predicated still runs counter to the individualistic focus of clinical practice. Perhaps, this historical legacy is one of the reasons that clinical trials often require the services of "outside" experts – such as statisticians and professional medical writers.

References

1. The Internet Classics Archive. On Airs, Waters, and Places by Hippocrates. 2016. [cited 2016 June 30]. Available from: http://classics.mit.edu/Hippocrates/airwatpl.html.
2. Collier R. Legumes, lemons and streptomycin: A short history of the clinical trial. CMAJ. 2009;180(1): 23-4.
3. Rusnock AA. Vital accounts: Quantifying health and population in eighteenth-century England and France. Cambridge: Cambridge University Press; 2002.
4. Louis PCA. Researchers on the effects of bloodletting in some inflammatory diseases, and on the Influence of tartarized antimony and vesication in pneumonitis. Putnam CG, translator. Boston: Hilliard, Gray; 1836.
5. Risueño d'Amador B. Memoire sur le calcul des probabilites applique a la medicine. Montpellier: L.F. Gaste; 1838.
6. Bulletin de l'Académie Royale de Médecine. 1836/37; 1: 702.
7. Gavarret J. Principes généraux de statistique médicale. Paris: Libraires de la Faculté de Médecine de Paris; 1840.
8. Tröhler U. Statistics and the British controversy about the effects of Joseph Lister's system of antisepsis for surgery, 1867-1890. JLL Bulletin: Commentaries on the history of treatment evaluation. 2016 [cited 2016 June 30]. Available from: http://www.jameslindlibrary.org/articles/statistics-and-the-british-controversy-about-the-effects-of-joseph-listers-system-of-antisepsis-for-surgery-1867-1890/.
9. Hardy A, Magnello ME. Statistical methods in epidemiology: Karl Pearson, Ronald Ross, Major Greenwood and Austin Bradford Hill, 1900-1945. Soziologie-Präventivmedizin. 2002; 47: 80–9.
10. Magnello E. The introduction of mathematical statistics into medical research: The roles of Karl Pearson, Major Greenwood, and Austin Bradford Hill. In: Magnello E, Hardy A, editors. The road to medical statistics. Amsterdam-New York: Rodopi; 2002. p. 95-123.
11. Matthews JR. Almroth Wright, vaccine therapy, and British biometrics: Disciplinary expertise versus statistical objectivity, In: Magnello E, Hardy A, editors. The road to medical statistics. Amsterdam-New York: Rodopi; 2002. p. 125-147.
12. Farewell V, Johnson T. Major Greenwood (1880–1949): A biographical and bibliographical study. Stat Med. 2016;35:645–70.

Author information

J. Rosser Matthews received his undergraduate degrees in mathematics and philosophy, and then earned a Ph.D. in the History of Science and Medicine. He is the author of the 1995 book *Quantification and the Quest for Medical Certainty*, which traced the history of statistical reasoning in 19th and 20th century medicine.