THE FOUNDATIONS OF EPIDEMIOLOGY: THE FRENCH CONNECTION*

David E. Lilienfeld and Abraham M. Lilienfeld, The Johns Hopkins University

We thought that we would begin our presentation of the historical foundations of both health statistics and epidemiology with a brief summary of the present state of both. Today, statisticians have several tools with which to handle data, some of which are shown in Table 1. These include the normal distribution, relative risk estimators, age-adjusted rates, confidence intervals, the census, experimental designs, and life tables. Yet all of these tools were developed and/or used by 19th century epidemiologists and statisticians during the historical period that I (D.E.L.) refer to as the Greening of Epidemiology.

The seeds of the epidemiological tree were sown by the French, shortly after the French Revolution of 1789 (1). Statisticians are familiar with the works of Laplace and Poisson in the early 1800's. Indeed, in his Philosophical Essay on Probability, Laplace describes our "modern" life table, makes the interesting statement that "A table of mortality is then a table of the probability of human life", and describes an approach to the analysis of competing risks (2).

The French Revolution represented a break with past traditions and customs (1). In medicine, a similar break with the past occurred. Prior to the Revolution, a theory of medicine provided the basis for its practice; afterwards, however, physicians decided to begin with the actual practice of medicine and generalized this into theory. The instrument of this generalization process was statistics, and the key figure in the medical adoption of statistics was the epidemiological pioneer, P.C.-A Louis. Louis, whom we view as one part of the "French Connection", championed what he called the "numerical method". He consistently employed this statistical approach to medicine, using it in 1830 to show the ineffectiveness of bloodletting as a therapeutic agent. Although Cochran has noted that modern experimental designs were used in agricultural research as early as the 1970's, it is interesting to note that Louis described a balanced block design in his book on the evaluation of bloodletting. He stated:

> "To this I reply that the calculus as I employ it, does not efface differences: it supposes them; it limits itself to combining similar unities in order to compare them with parallel unities, these being subjected to somewhat different influences; that is, after all, as has been before remarked, it should sometimes be necessary that facts should be combined which are not strictly similar. The error will be distributed through the different groups or classes of facts, and will be equalized; so that a comparison can be instituted between several groups without altering the result."(3).

Louis was also familiar with the general concept of a prospective study, as shown in his approach to the determination of whether or not phthisis was inherited, stating:

> "...to determine the question satisfactorily, tables of mortality (life tables) would be necessary, comparing an equal number of persons born of phthistical parents with those in an oposite condition.(4).

One of the ideas that dominated 19th-century epidemiology and statistics was the concept of a "law of mortality" or "vitality". These "laws" were not necessarily mathematical, such as Gauss' Law, although they could be; one such law was the doctrine of contagium vivum, the "germ theory" Louis was an advocate of the idea of a law of mortality, as shown in a letter to James Jackson, the father of one of his students, in which he wrote:

> "Think for a moment, sir, of the situation in which we physicians are placed. We have no legislative chambers to enact laws for us. We are our own lawgivers' or rather we must discover the laws on which our profession rests. We must discover the laws and not invent them; for the laws of nature are not to be invented" (5).

Thus, it can be seen that statistics was the quantitative manifestation of the inductive reasoning used by the Parisian school of medicine, following the pattern established by the physicists, who were using the calculus as a quantitative manifestation of the deductive reasoning process for deriving physical laws. Time does not permit a presentation of all of Louis' contributions to statistics and epidemiology.

In 1840, Jules Gavarret, of the Polytechnic School of Paris, a student of Quetelet, published a book entitled "General Principles of Medical Statistics", in which 99% confidence intervals were applied to Louis' data on the effect of bloodletting (6). Gavarret criticized Louis because the latter would not use such confidence intervals. Perhaps Louis understood the distinction between biological and statistical significance, when used in epidemiology better than we do today!

Louis' influence was extended by the work done by his students, among whom were the leaders of mid- and late-19th century epidemiology and statistics. These students form the second part of the French Connection. The American students of Louis are shown in Figure 1. Louis' European students are shown in Figure 2. One of these students, familiar to everyone here, was William Farr.

Farr, one of the leading statisticians of his time, was a titan of mid-19th century epidemiology. As Louis's student in the early 1830's he was instilled with Louis' beliefs in a "law of mortality", having stated:

> "Thus, we learn in the same circumstances the same number of people die at the same ages of the same diseases, year after year; organized bodies governed by laws as fixed as those which govern the stars in their courses" (7).

And, further:

> "The deaths and causes of death are scientific facts which admit of numerical analysis; and science has nothing to offer more inviting than the laws of vitality..." (7).

Throughout his life, Farr searched for these laws by constructing life-tables, etc. For it was Farr who referred to the life table as a "biometer" because of its ability to measure life (1). It appears that either Farr or one of his fellow actuarians, in the late 1830's or early 1840's, coined the term "force of mortality" as used in life tables. One should be aware that this occurred in the mid-1800's when the physical laws governing electrical forces were being discovered, when physicists were using such terms as the "electromotive force". It is possible that the "force of mortality", was therefore used in a similar manner in various laws of vitality.

(1,7) Farr had an amazing grasp of epidemiologic concepts, as shown in Table 2, and, of course, Farr's guiding philosophy is shown by his statement "The death rate is a fact; anything beyond this is an inference"(7). He worked with one of Louis' other students, Marc d' Espine, to develop the predecessor of today's International Classification of Diseases. Lastly, Farr may be viewed as one of the founders of the English school of statistics. He was active in the Statistical Society of London, predecessor of the present Royal Statistical Society, eventually becoming its president in 1872. He established the British Vital Registration System, including the first turly "modern" national census. He had a strong influence on Francis Galton, apparently interesting Galton in statistics. And, he was an advocate of the rigorous statistical analysis of epidemic data; indeed, in 1854, he noted the relationship of water purity, by water company, with cholera mortality in a statistical manner, laying the foundation for John Snow's classic analysis of the epidemic (Table 3).

William A. Guy, a physician, was another student of Louis' at a time after Farr had already returned to London. He, along with Farr, was one of the founders of the London School of Statistics. He became a Fellow of the Statistical Society of London, and was very active in its activities, serving as editor of its journal and in 1874, as its president. In 1846, Guy became the Dean of the King's College Medical School. He was among the first to note the basis of what today we term a "Berksonian Bias". He stated:

> "There are two questions to which I am not aware that any answer has yet been given; nor has any collection of facts been made with a view to furnish a reply. The first question refers to the class of persons who resort to hospitals; the second to the proportion which that class forms of the population to which they belong" (8).

Guy also noted that the variation of the estimator of the mean in a sample decreased with an increasing sample size (9).

The appropriateness of any application of Gavarret's and other French statistician's theories of probability and statistics to clinical medicine and epidemiology was also investigated by Guy. Unlike today's statisticians and epidemiologists, I guess that Guy did not believe that a statistical <u>theory necessarily</u> had any relevance to the "real world". Indeed, the modern multiple logistic equation, the present fetish of both of both statisticians and epidemiologists, has advanced epidemiological research to such an extent that figure 3 shows our present predicament!! In 1855, Guy stated:

> "Gavarret...criticises with some severity the conclusions of Louis respecting pulmonary consumption and fever, on the score of the insufficient number of his facts (collected over 7 years), and insists on applying to those conclusions corrections avowedly drawn from treatises on the doctrine of probabilities. Now, unless I am greatly mistaken, no attempt of any kind has yet been made to show that rules and calculations derived from abstract reasoning upon probabilities, backed by a few experiments on occurrences brought about by what is commonly designated 'chance', are applicable to events of a totally different order, brought about by the operation of the human will or by the multitudeinous external influences which, acting on the human frame, preserve it in health or give rise to the diseases which impair its vigour and utlimately destroy it" (10).

Accordingly, Guy proceeded to perform a series of experiments which would be collectively known today as a "Monte-Carlo" simulation.

He began with a defined population of black and white balls; he proceeded to randomly sample from that population, both with and without replacement. Guy then analyzed the samples and compared the results with the population. He concluded that there was some analogy with Gavarret's theorems and statistical significance. Strangely, even after this simulation, Guy used such methods only once.

Lastly, one of Guy's great contributions to epidemiology was his epidemiological studies of the effect of occupation on health. They read as if they were reports of studies published in one of today's scientific journals. Indeed, in one of Guy's studies, a numerical expression equivalent to the odds ratio was used, seemingly, as an estimate of relative risk, which was well-known to have been used by mid-19th century epidemiologists and statisticians. One other important figure was an actuarian, F.G.P. Neison, a colleague of Guy's occupational studies. Neison was the <u>first</u> person to use a method of standardization for death rates to account for differences in the age distribution of populations (Table 4) (11).

Of course, after Galton, the history of statistics is well-known. But, there are some recently uncovered details such as the relationship of Guy to Newsholme, who wrote a text entitled, "Elements of Vital Statistics" in 1889 (12). The tree of epidemiology has grown quite a bit since it first "greened" in the mid-1800's, when epidemiology was barely distinguishable from statistics. Many new branches have grown on that tree as epidemiology has developed. Yet, unfortunately, as a result of many circumstances, we believe that we, today's statisticians and epidemiologists, are out on a limb of this epidemiological tree, and that that limb is being cut off from the trunk! One reason for this, we believe, is indicated in Figure 4.

Thus, Kendall's question, posed in 1975, of why statistics developed in the way that it did has been partially answered (13). Statistics was the means of generalization in Post-Revolutionary French medicine. There were two approaches to such generalizations, known as laws of mortality.

Louis advocated a discrete approach - an absolute law - and, hence, had little real use for theoretical statistics; Gavarret advocated a stochastic approach - a probabilistic law - and, hence, freely used theoretical statistics. Today, these two approaches are evident in physics in the form of the classic Einstein-Heisenberg controversy, in epidemiology with the "web of causation" and definite specific causes debate and, generally, in deterministic and stochastic equations.

The greatest value of history is the perspective it allows one to view the present, before it, too, becomes history. One lesson that we have learned from our on-going historical excursions is that the basic structure of epidemiology is composed of methods - methods devised by the epidemiologist and the statistician alike. These methods should continue to be developed by both the epidemiologist and the statistician almost hand-in-hand. The inferences derived from any given study can change, but the method used to conduct that study does not. Indeed, one reason why the histories of both epidemiology and health statistics have not yet been written is the over-emphasis on inferences and the lack of attention to methods. For, as Daniel C. Gilman said of the Johns Hopkins University in 1890:

> "Whatever gains we may make in our material condition, whatever limitations are still obvious, let us not forget , my friends, that men and methods make universities, not halls, nor books, nor instruments, important as these are."

So, the same can be said for epidemiology:

Whatever limitations are still obvious, let us not forget that men and methods make epidemiology, not statistical significance levels, nor computers, nor inferences, important as these are.

---

*Partially supported by a grant from the Milbank Memorial Fund.

References

1) Lilienfeld DE: The greening of epidemiology, sanitary physicians and the London Epidemiological Society (1830-1870). Bull Hist Med (In press)
2) Laplace PS: A Philosophical Essay on Probabilities. New York, Dover Publications, 1951
3) Louis PCA: Researches on the Effects of Bloodletting in some Inflammatory Diseases, and on the Influence of Tortarized Antimony and Vesication in Pneumonitis. Translated by C.G. Putham. Boston, Hilliard, Gray and Co. 1836
4) Louis PCA: Review of pathological researches on phthisis. Amer J Med Sci 19:445-449, 1836
5) Bollet AJ: Pierre Louis: The numerical method and the foundation of quantitative medicine. Amer J Med Sci 266:92-101, 1973
6) Gavarret J: Principles Généraux De Statstique Médicale. Paris, Bechet Jeune et Labe, 1840
7) Farr W: Vital Statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr. Metushen, NJ, New York Academy of Medicine, 1975
8) Guy WA: On the nature and extent of the benefits conferred by hospitals on the working classes and the poor. J Roy Stat Soc 19:12-27, 1856
9) Guy WA: On the relative value of averages derived from different number of observations. J Roy Stat Soc 13:30-45, 1850
10) Guy WA: On the analogy between the aggregate effects of operation of the human will and the results commonly attributed to chance. J Inst Act 5:315-323, 1855
11) Neison FGP: On a method recently proposed for conducting inquiries into the comparative sanatory condition of various districts. J Stat Soc 7:40-68, 1844
12) Newsholme A: Elements of Vital Statistics. New York, The Macmillan Co., 1899
13) Kendall MG: Statisticians - Production and consumption. Amer Statistician 30:49-53,1975

TABLE 1

TOOLS OF THE MODERN HEALTH STATISTICAN AND EPIDEMIOLOGIST

1) THE NORMAL DISTRIBUTION
2) RELATIVE RISK ESTIMATORS
3) AGE-ADJUSTED RATES (MORTALITY, MORBIDITY, ETC.)
4) CONFIDENCE INTERVALS
5) THE CENSUS
6) EXPERIMENTAL DESIGNS (CROSS-OVER, LATIN SQUARE, ETC.)
7) LIFE TABLES

TABLE 2

Examples of William Farr's Understanding of Epidemiologic Concepts

Epidemiologic concept	Farr's statement
Scope of epidemiology	"The causes that make the rates of mortality vary may be considered under two heads - (1) Causes inherent in the population itself, such, for example, as sex and age. (2) Causes outside the population, such as air, water, food, clothing, dwellings, or such groups of causes as are involved in residence, and relation of the several parts to each other in time and space."
Person-years	"A year of life is the lifetime unit. It is represented by one person living through a year; or by two persons living through half a year."
Relationship of death rate and probability of dying (or living)	"...the rate of mortality serves to give the probability of living a year..."
Standardized mortality rate	"[If] the number of boys under 5 years of age was 147,390; the annual rate of mortality in the healthy districts [the standard population] was .04348;...6367 deaths which would have happened in London... continuing the process... the mortality in London should [be] 15 in 1,000..."
Dose-response effect	"...the effects are in some regulated proportion to the intensity of the causes...
Need for large numbers of population and biological inferences	"...When the number of cases is considerable the relative mortality is most correctly expressed and...slight differences deserve little attention."
Herd immunity	"The small-pox would be...sometimes arrested, by vaccination which protected a part of the population..."
Prevalence = incidence X duration	"...in estimating the prevalence of disease, two things must be distinctly considered; the relative frequency of their attacks, and the relative proportion of sick-time they produce. The first may be determined at once, by a comparison of the number of attacks with the numbers living; the second by enumerating several times the living and the actually sick of each disease, and thence deducing the mean proportion suffering constantly. Time is here taken into account: and the sick-time, if the attacks of two diseases be equal, will vary as their duration varies, and whatever the number of attacks may be, multiplying them by the mean duration of each disease will give the sick-time."
Retrospective and prospective studies	"Is your inquiry to be retrospective or prospective? If the former the replies will be general, vague, and I fear of little value..."

TABLE 3

Mortality from Cholera in Districts Supplied by Water Companies, 1853

		Aggregate of Districts Supplied Chiefly by the Respective Water Companies			
Water Companies	Source of Supply	Elevation (in feet) above trinity high water mark	Population	Deaths from cholera in 13 weeks ending Nov. 19	Deaths per 100,000 inhabitants
London	-	39	2,362,236	744	30
Hampstead & New River	Springs at Hampstead and Kenwood, two artesian wells and New River	80	166,956	8	5
New River	Chadwell Springs in Hertsforshire, from River Lee, and four wells in Middlesex and Herts	76	634,468	56	9
Grand Junction	Thames, 360 yards above Kew-Bridge	38	109,636	15	15
Chelsea	Thames at Battersea	7	122,147	22	18
Kent	Ravensbourne in Kent	18	134,200	31	23
West Middlesex	Thames at Barnes	72	277,700	89	32
East London	Lee at Lee Bridge	26	434,694	162	37
Lambeth & Southwark	Thames at Thames Ditton and at Battersea	1	346,363	220	64
Southwark	Thames at Battersea	8	118,267	121	102
Southwark & Kent	Thames at Battersea, Ravensbourne in Kent, Ditches and wells	0	17,805	19	107

TABLE 4

CHADWICK'S CALCULATIONS

	BETHNAL GREEN	MARYLEBONE
MEAN AGE AT DEATH (UNADJUSTED FOR AGE)	25.8	29.12

NEISON'S CALCULATIONS

	BETHNAL GREEN	MARYLEBONE
MEAN AGE AT DEATH (ADJUSTED FOR AGE)	25.8	24.52

Figure 1: The influence of P-C. A. Louis on the development of statistics and epidemiology in the nineteenth century in the United States.

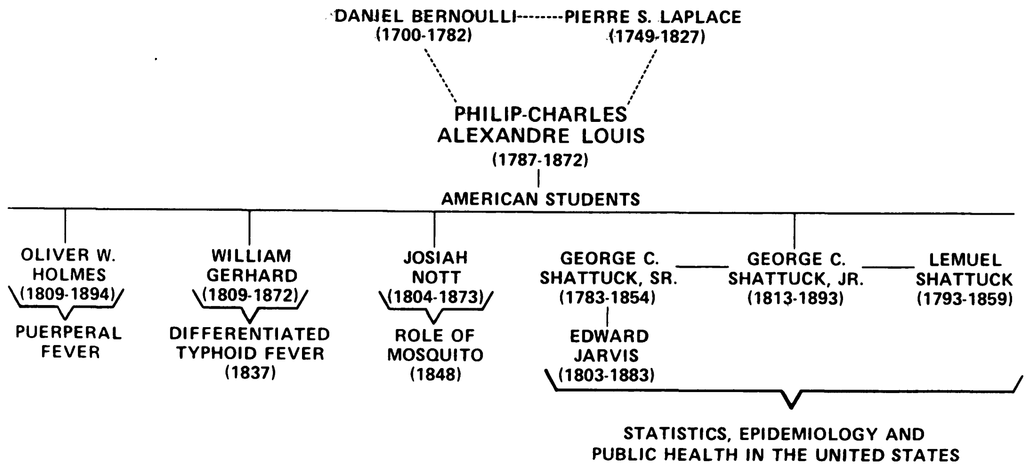
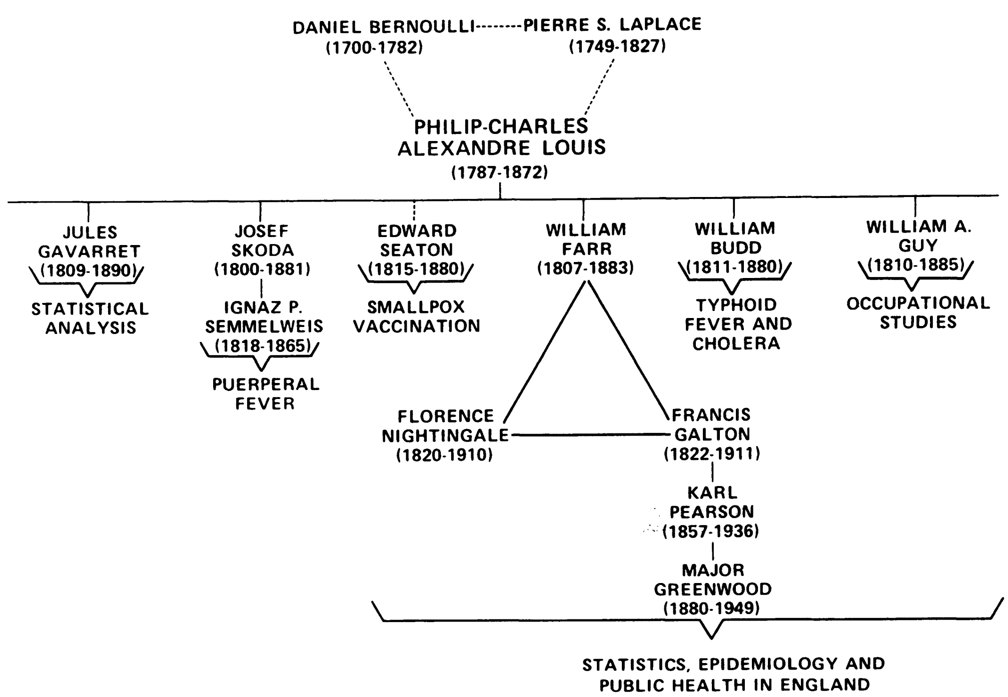
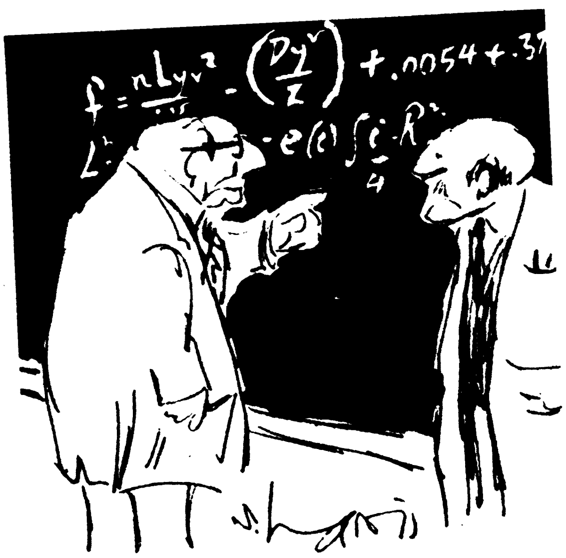


Figure 2: The influence of P-C. A. Louis on the development of statistics and epidemiology in the nineteenth century in England and Europe.





"Does this apply always, sometimes, or never?"

". . . and in 1/10,000th of a second, it can compound the programmer's error 87,500 times!"