The evolution of the causation concept and its relation with statistical methods in Epidemiology

Evolução do conceito de causa e sua relação com os métodos estatísticos em Epidemiologia

Luis Fernando Lisboa*

ABSTRACT

A historical review places the first registers of Epidemiology in ancient Greece, with Hippocrates, who identified environmental causes of diseases. Along the centuries, the evolution of the causation concept started to be related to changes in scientific paradigms. In London, during the 17th century, the quantitative method was introduced in Epidemiology, but it was only by the end of the 19th century that the concept of the environment and a mathematical approach to understanding Public Health issues were well established. This was a very rich period to setting new concepts and systematizations in epidemiologic methodology. The beginning of the 20th century consolidated Epidemiology as a scientific discipline and the development of computers in the post-war years brought much advance in this field. Nowadays, Epidemiology plays an important role as it integrates scientific knowledge on the health/disease process to the professional area, participating in population healthcare efforts.

Keywords: Epidemiology/history; Epidemiology/statistics & numerical data; Uses of epidemiology

RESUMO

Uma análise histórica situa os primeiros registros da Epidemiologia na Grécia antiga com Hipócrates que pretendia atribuir causas ambientais às doenças. Ao longo dos séculos, a evolução do conceito de causalidade passou a se relacionar com mudanças no paradigma científico. No século 17, em Londres, o método quantitativo foi introduzido na Epidemiologia, mas apenas no final do século 19 os conceitos de ambiente e de abordagem numérica na compreensão dos problemas de Saúde Pública ficaram bem sedimentados. Esse foi um período muito rico no estabelecimento de novos conceitos e sistematizações na metodologia epidemiológica. No início do século 20, a Epidemiologia foi consolidada como disciplina científica, e o aparecimento dos computadores, no período do pós-guerra, trouxe grande desenvolvimento para a área. Atualmente, a Epidemiologia ocupa um papel de destaque na integração dos conhecimentos científicos sobre o processo saúde/doença ao campo profissional, desenvolvendo esforços pelo cuidado da saúde das populações.

Descritores: Epidemiologia/história; Epidemiologia/estatística & dados numéricos; Aplicações da epidemiologia

According to Last, we may conceptually define Epidemiology as "the study of the frequency, distribution and determinants of states or events related to health of specific populations, and the application of these studies to control health problems"[1]. Among the various terms in this definition, we highlight determinants, which embodies the main purpose of Epidemiology: to seek the cause and factors that influence the occurrence of events related to the health process[2].

Since the beginning of time, man has tried to understand what cause something to exist, what are the determining factors of an event, the origin and the links that relate phenomena and establish that one, or many, is a basic condition for the existence of others; in other words, there is an interest the investigation of causes.

A cause may be understood as any event, condition or feature that is functionally essential for a disease to occur[2].

The concept of causation developed together with a change in the paradigm of scientific knowledge towards a strong empirical observation component, which drove forward an epidemiological approach and statistical methods.

The first writings about Epidemiology may be found in Ancient Greece (400 B.C.), when Hippocrates disagrees with the idea of a divine origin of disease, in a classical work named "On airs, waters and places," and discusses its environmental causes. He suggests that the local climate, water or favorable winds of a place are elements that help physicians in the evaluation of the general health conditions of the inhabitants of a given place; such thought raise basic epidemiological

*Post-graduate student in Public Health at the Universidade de São Paulo – USP, São Paulo (SP), Brazil.

Corresponding author: Luis Fernando Lisboa – Rua Jaguaribe 760 – apto. 21 – Vila Buarque – CEP 01224-000 – São Paulo (SP), Brasil – Tel.: 11 3663-6343 – e-mail: luislisboa@einstein.br

Received on Jan 29, 2008 – Accepted on Jul 17, 2008

concepts, such as the environment (represented by air, water, place) and the host (represented by the individual configuration)[3]. Other works, such as "The book of prognostics and aphorisms", anticipated the then revolutionary idea that physicians could predict the progression of a disease by observing a sufficient number of cases. This theoretical structure proposed by Hippocrates was kept alive by his successors, such as Galen (201 b.C.), who believed that each disease was connected to specific organs, and stated that the health of man depended on four basic humors (blood, yellow bile, black bile and phlegm)[4] being balanced.

The miasmatic theory of disease was a curious theory that persisted for many years; it suggested that vapors or miasmas issued from certain soil types (particularly swamps) or from night airs could cause diseases in people that came into contact with them. In Shakespeare's Julius Caesar, Portia warns her husband Brutus against the ills of walking out in the middle of the night:

> "Is Brutus sick? and is it physical
> To walk unbraced and suck up the humours
> Of the dank morning? What, is Brutus sick,
> And will he steal out of his wholesome bed
> To dare the vile contagion of the night,
> And tempt the rheumy and unpurged air
> To add unto his sickness?"[5]

There was not yet, however, a collective notion of the health-disease process[6]. It is not by chance that the concepts of population, State, collectiveness arose in the beginning of 17th century, initial period of the Modern Age in which the capitalist mode of production was predominant, what would transform society. John Graunt's (1620-1674, London) noteworthy work introduced quantitative methods in Epidemiology with his treatise on mortality tables. Graunt used mortality reports from parishes to establish gender, urban/rural and timeline differences in mortality[3].

In Europe, in the mid 19th century, amidst the Industrial Revolution, with populations moving into cities, and with cholera, typhoid fever and yellow fever epidemics, experts were still divided between the miasmatic and the germ theories. Towards the end of the 19th century, however, the concept of the environment and a numerical approach to understanding Public Health issues were well established in epidemiological thinking, thanks to contributions from pioneers such as the French Pierre Louis (1788-1872). He worked with data of hospitalized patients to introducing statistical methods for counting events, and related the lethality of pneumonia to the moment that bloodletting was started. The French Louis Villermé (1782-1863) investigated poverty and working conditions, their effects on health and the close relation of social and economic conditions to mortality (for instance, the health of workers in the cotton, wool and silk industries). The British William Farr (1807-1883) worked during 40 years in the General Register Office in England and became involved in classifying diseases and describing the laws of epidemics (Farr's law). He is considered the father of vital statistics and surveillance.

The most significant contribution to Epidemiology in this period is attributed to John Snow because of his treatise "On the mode of transmission of cholera," published in 1855, which contains his observations about cholera in London, between 1849 and 1854. He described, in a logical sequence, the behavior of cholera based on mortality data by studying the frequency and distribution of deaths according to chronology of facts and site; he also surveyed other factors related to the cholera cases to make causal hypotheses. Although Snow's ideas developed under Morgagni's and Henle's concepts, his treatise introduced new concepts, such as cause and effect, the measure of disease occurrence, contrast and effect measures and control of systematic errors (biases), and systematizes epidemiological methodology, which remained nearly unchanged until the mid of the 20th century[7]. Snow's epidemiological reasoning may be summarized as follows:

1. description of the behavior of cholera according to time, state and person;
2. search for causal associations between diseases and specific factors by investigating the facts; assessment of existing hypotheses; formulation of more specific hypotheses; collecting additional data to test new hypotheses[7].

Still in the 19th century, the French Louis Pasteur (1822-1895) established the biological basis of infectious diseases and introduced another important epidemiological concept: host resistance and immunity[3]. Robert Koch (1843-1910), also French, was one of the founders of microbiology (together with Pasteur) and mainly responsible for the current understanding of the Epidemiology of communicable diseases; he also discovered the causal relation between *M. tuberculosis* and the disease and created what today is referred to as the 'Henle-Koch postulates', which emphasize the infectious etiology of diseases, and are listed below:

1. the microorganism must be found in abundance in all organisms suffering from the disease;
2. the microorganism must be isolated from a diseased organism and grown in pure culture;
3. the cultured microorganism should cause disease when introduced into a healthy organism;

4. the microorganism must be re-isolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.

Under the influence of microbiology, studies were done in laboratories and the other areas of medicine became subordinated to this knowledge. Public health specialists were trained in laboratories and Epidemiology started to be used in the study of non-infectious diseases. A highlight is the work done by Joseph Goldberger, who established the etiology of pellagra regarding deficiency; stratification of the analysis was done for controlling its effects.

Epidemiology was consolidated in the beginning of the 20th century as a scientific discipline when many Universities worldwide created departments of Epidemiology. Wade Hampton Frost headed the Epidemiology department at the Johns Hopkins University and established the basis of cohort studies and survival analysis as a technique for data analysis. In 1947, Austin Bradford-Hill and Richard Doll led a historical case-control study at the London School of Hygiene and Tropical Medicine, establishing the relation between smoking and lung cancer. Austin Bradford-Hill is also considered as responsible for introducing patient randomization in clinical trials, a reference for case-control studies due to its methodological rigor. These authors did not calculate the effects as Odds Ratio, which would later be described as a "method for estimating comparison ratios" by Jerome Cornfield in 1951. This was a significant contribution to strategies in data analysis of case-control studies.

Epidemiological methodology advanced further with the incorporation of statistics in the post-war period, stimulated by computers – which were just then being developed – and dealing with a broad spectrum of health conditions. Studies by Doll and Hill, and the investigations of cardiovascular disease conducted in the city of Framingham, USA, are two examples of the use of epidemiological methods in chronic diseases[7].

In 1950, John Gordon published a paper entitled "Epidemiology: old and new", in which he discusses the aging of populations. In 1956, Gordon, with other authors, published the article "The community problem in coronary heart disease: a challenge for epidemiological research", providing an optimistic view about the instrumental role that knowledge about Epidemiology could have in investigating chronic-degenerative disease.

This transition from infectious to chronic-degenerative diseases led to an expansion in the concept of causation, which went from a monocausation model to what could be referred to as the 'causation network', wherein the concept of an etiological cause gives way to the concept of predisposing factors or disease risk[8]. The first reference to the causation network appeared in 1960, in the textbook by MacMahon e Pugh entitled "Epidemiology: principles and methods", which classified and organized the conceptual and methodological development of Epidemiology.

In 1965, Hill suggested nine criteria (or association aspects) to be considered when separating a causal from a non-causal association[9]:

1. strength;
2. consistency;
3. specificity;
4. temporality;
5. biological gradient;
6. plausibility;
7. coherence;
8. experimental evidence;
9. analogy.

The comparison between Koch's and Hill's criteria shows the progression of different perspectives about the health-disease process; in the first there is expectation of monocausation while in the latter there is the specificity between cause and effect.

In the succeeding years, the main conceptual issue in Epidemiology was the interaction between causation factors, which underlined the concept of causation network. A historical reference is the investigation of interactions between oral contraceptives and arterial hypertension and their relation with the occurrence of thromboembolic phenomena. Marshall and Warren studied the association between peptic ulcers and the *Helicobacter pylori*, which is an example of the failure of infection/chronic nomenclature. Further studies are the many recent trials on the genetics of cancer, which are an example of the challenge of reviewing the etiological concept of disease.

In 1985, Miettinen proposed the idea of 'occurrence function' to describe the relation between cause and effect, based on generalized linear models for the statistical analysis of associations, such as logistic regression, which is much used. Rothman, however, brings attention to the failure of additive or multiplicative models, since selection of an n-order interaction model may be subjective and may help judge data adjustment, but cannot provide inferences beyond the sample universe. To overcome these difficulties, Rothman proposed methods for calculating effects by logical propositions and intersections of sets of causal factors. His general causation model suggests that, for a disease to occur, a set of component causes is required. We may succinctly define sufficient cause as the minimum set of events and conditions that will inevitably result

in a given disease; when certain component causes are present in all of the alternative sufficient causes, they are named necessary causes[10].

Nowadays, we have the concept of process of causal relations, which provides an adequate treatment of interactions within cause processes. Krause stated that "the notions of cause and effect, and determinism in general, have lost much of their importance, since the advent of quantum mechanics, especially bearing in mind its probabilistic interpretation by M. Born"[11], although at the same time he states that the notion of causation is indispensable in Medicine.

Studies carried out by the Canadian epidemiologist Pierre Philippe suggested that the chaos theory may be an alternative model. In 1992, Philippe tested the adjustment of different models to a description of measles in New York throughout many decades, and found that an auto-regression chaotic model would best fit the data[12].

Current investigation of causal relations is done at a molecular level; according to Susser, we are witnessing the birth of a new Epidemiology[13]. Recently, knowledge and analytical alternatives have been employed, including Bayesian methods, fuzzy logic, fractal systems and neural networks.

Epidemiology has the unique role of conciliating its roles as a scientific discipline, as a method for developing original knowledge about the health-disease process and also participating in population healthcare efforts. We should be ready for paradigm changes and for facing new challenges with open arms. As Sir Isaac Newton said, "What we know is a drop, what we do not know is an ocean".

REFERENCES

1. Waldman EA. A epidemiologia em medicina In: Lopes AC, Amato NV. Tratado de clínica médica. 2a. ed. Volume 1. São Paulo: Roca; 2007.
2. Medronho R. Epidemiologia. Rio de Janeiro: Atheneu; 2003.
3. Guilam MC. O conceito de risco: sua utilização pela epidemiologia, engenharia e ciências sociais [Internet]. [cited 2008 Abr 9]. Disponível em: http://www.rnsp.fiocruz.br/projetos/esterisco.
4. Freezer LW. Theories concerning the causation of disease. Am J Public Health [Internet]. 1921 [cited 2008 Sep11]; (10):908-12. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1353971
5. Shakespeare W. Júlio Cesar [Internet] [cited 2008 May 30]. Disponível em: http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=2345.
6. Barata RB. Epidemiologia e saber científico. Rev Bras Epidemiol. [Internet] 1998 [cited 2008 Ago 22]; 1(1):14-27. Disponível em: http://www.scielo.br/pdf/rbepid/v1n1/03.pdf
7. Waldman EA. Vigilância em saúde pública, volume 7. São Paulo: Faculdade de Saúde Pública da Universidade de São Paulo; 1998. (Série Saúde e Cidadania)
8. Pereira JCR. Antecedentes históricos, conceitos e usos da epidemiologia [notas de aula]. São Paulo: Faculdade de Saúde Publica; 2007.
9 Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965;58:295-300.
10. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95 Suppl 1:S144-50.
11. Krause D. A filosofia da ciência de Newton C.A. da Costa. São Paulo: IEA-USP; 2001. Série Lógica e Teoria da Ciência.
12. Philippe P. Chaos and public health: implications for an epidemic. Can J Public Health.1992;83(2):165-6.
13. Susser M, Susser E. Choosing a future for epidemiology: I. Eras and paradigms. Am J Public Health. 1996;86(5):668-73.