

Causality in epidemiology

Summary

Epidemiology represents an interesting and unique example of cross-fertilization between social and natural sciences. Epidemiology has evolved from a monocausal to a multicausal concept of the “web of causation”, thus mimicking a similar and much earlier shift in the social sciences. However, in comparison with the social sciences epidemiology is both more sensitive to underlying biological models (which condition the interpretation of population findings), and more prone to a simplification of the causal pathways. Paradoxically, epidemiology has developed more sophisticated theoretical models for bias and confounding than the social sciences did, but for the practical purpose of identifying single preventable risk factors. Epidemiology makes use more often of study designs that simulate experimentation, than of surveys in the general population.

Keywords: Probabilistic models – Proof – Experimental approach – Conditional counterfactuals – Conditionalized realism – Social sciences.

Causality in medicine can be viewed from at least three perspectives. One refers to the classical epistemological question of how causes are discovered and which is the most effective model of explanation. This perspective includes for example the concept of the “arc of knowledge”, i.e., the respective roles of hypothesis generation and hypothesis testing (Odlroyd 1986). The second perspective refers to the “burden of proof” which is needed to consider an agent as a cause of disease, i.e., how much evidence should be collected before practical action is justified. The story of Semmelweis is exemplar from both perspectives: from the epistemological point of view, the work of Semmelweis, which is summarised below, has been considered as paradigmatic by Carl Hempel (1966). In the meantime, according to a popu-

lar (but inaccurate) reading of his life, the resistance to accept the practical implications of that work has been considered as an example of the conflict between academic interests, scientific standards and ethical obligations of the profession. In fact, a romantic interpretation of the relationships between Semmelweis’ genius and the obtuseness of his time’s academic community is mainly the invention of the writer Ferdinand Céline and is not supported by scholarly work (see e.g., Nuland 1989).

The third perspective is related to how we interpret the relationship between cause and effect, for example according to a probabilistic or a deterministic model (the latter implying that the effect is predictable from the cause at the individual level).

Some antecedents of the modern notion of cause in medicine

The heritage of Claude Bernard

The work of Claude Bernard has played a key role in last century’s medicine. I will just sketch some of his main contributions that are relevant to conceptions of causality. His well-known theory of “milieu intérieur” (internal medium) has been instrumental to the discoveries of endocrinology and, more generally, to the identification of regulatory mechanisms in physiology. The underlying idea was that the autonomy of the living being from the external world is due to the ability to regulate his internal milieu and to keep it in equilibrium in spite of environmental changes (for example, rates of glucose in the blood are kept constant in spite of variable dietary intake). This idea of “homeostatic” mechanisms, based on a feed-back between the environment and the living being, has been applied to several aspects of physiology and pathology, and is a key issue in causality. For example, the current theory of carcinogenesis is based on the central role played by DNA (see below) and the related

hypothesis of a “homeostatic” mechanism of DNA repair: cancer arises because the balance between carcinogenic “hits” and DNA repair is overcome.

But Bernard’s work is important also from the epistemological point of view. According to his book “Introduction to experimental medicine” (Bernard 1957), theories are not built on the basis of cumulation of observations, like in the empiricist account of knowledge. Rather, there is a feedback mechanism between facts and theories, a “tension” between the proliferation of new observations and their ability to fit existing theories (like in a homeostatic model).

Virchow’s school

Because of its relevance to the questions we are discussing, a debate which took place in medicine around 1900 is worth describing. I refer to the controversy between Rudolf Virchow’s cellular pathology on one side and bacteriology on the other side, also referred to as the controversy between *aetiology* and *pathology* (von Engelhardt 1993). Klebs, a student of Virchow, denied “the autonomy of the cell as a principle of disease” (i. e., the internal origin of diseases), and insisted on external causes of human pathology. Later, he refined his theory by claiming that infectious diseases can be explained neither solely from the viewpoint of the bacteriologist, nor from that of the cellular pathologist; rather, diseases represented a battle between bacteria and cells (von Engelhardt 1993).

Subsequently, several authors, in particular Martius and Hueppe, insisted that man has to free himself from a one-sided etiological view of thinking, which aims to recognise a single and necessary cause for each event. Following the functionalist philosophy of Mach, Martius interpreted “disposition” (internal cause) and “stimulus” (external cause) as variables subject to mathematical treatment (von Engelhardt 1993).

According to the more radical school of “conditionalists”, the word “cause” should be completely substituted by the concept of “determining condition”, where “all the conditions of a process or state are equally important”, i. e., a hierarchy cannot be established. In the opinion of von Hase-mann, the monocausal viewpoint has been promoted in medicine because of a practical interest in therapy, but from a scientific perspective monocausality must be opposed. Conditionalism stimulated the idea of a multiplicity of etiological factors and insisted on the strict relationship between external determinants and internal reaction to the stimulus. From a practical point of view, conditionalism was not able to ascribe relative importance to the determinants, and it failed in the court and in insurance decisions.

The relevance of this debate to contemporary epidemiology is rather clear. For example, O. S. Miettinen has proposed to

treat causal relationships in medicine according to the concept of “occurrence function” (Miettinen 1985); such approach is very similar to the mathematical treatment by Martius. The occurrence function is

$$\text{disease} = a + \text{determinant(s)} + \text{confounder(s)} \\ + \text{effect modifier(s)}$$

i. e., the probability of disease is explained by the conjunction of one or more determinants, after allowing for confounding variables and for other exposures that “modify the effect” of the determinant. For example, cigarette smoking is a determinant of lung cancer, after allowing for occupational exposure as a potential confounder and considering consumption of Vitamin C as an effect modifier.

Hempel on Semmelweis

Carl Gustav Hempel, in one of his more important texts on science philosophy (Hempel 1966), describes the work of Ignatz Semmelweis as an example of correct scientific procedure based on trial and error. Semmelweis described the occurrence of “puerperal fever” in the course of time and in two different obstetric clinics of the Vienna General Hospital: in the first period (1833–1858) mortality from puerperal fever was very similar in the two clinics, while in the second (1840–1846) it was remarkably different: in fact, 1989 deaths occurred among 20042 women in the first clinic vs. 691 among 17791 in the second. The only element that could differentiate the two time periods was that in 1840–1846 the students attending the first clinic were involved in autopsies as a part of their training. This suggested to Semmelweis the idea that the cause of puerperal fever could be related to some “particle” transmitted from the corpses; this idea was reinforced by the death of Kolletschka, a professor of pathology and friend of Semmelweis, who apparently died from a disease very similar to puerperal fever after hurting his finger accidentally during an autopsy (but the distinction between reality and legend is not clear-cut). After 1846, Semmelweis made an experiment, by introducing the very simple practice of washing hands after autopsies and before visiting the women in the obstetric department. Mortality from puerperal fever declined abruptly to 3% or less in both clinics.

Hempel (1966) describes the inferential procedures used by Semmelweis as an example of “trial and error”: after the early epidemiological observations suggesting transmission of “particles” from corpses, he looked for falsifying evidence and alternative explanations. For example, he ruled out that the disease was contracted before hospitalisation for causes related to the women’s living conditions. He also ruled out, partly by conducting experiments, several other potential

confounders or sources of bias. Therefore, the work of Semmelweis can be described according to the idea of the “*arc of knowledge*”, i.e., research can be summarised into four phases:

1. hypothesis generation from clinical observations
2. reinforcement of the hypothesis on the basis of planned epidemiological observation
3. testing of the hypothesis with a formal study design (in clinical research the gold standard for this phase is represented by the randomised clinical trial)
4. intervention, i.e., deliberate change of exposure circumstances in order to prevent disease occurrence.

Phase 1 and 2 belong to the ascending part of the arc of knowledge (induction), while parts 3 and 4 belong to the descending arc (deduction).

The logical reconstruction of Semmelweis’ work proposed by Hempel (1966) is a nice introduction to causal thinking in a non-experimental science like epidemiology, though one may not completely agree with the “trial-and-error” model he proposes. In contrast e.g., with Popper, Hempel does not stress the limitations of the observational (non-experimental) nature of epidemiology. (Incidentally, the description of Semmelweis’ work cannot be separated by his biography, and in particular by his mental illness and tendencies to self-destruction; see Nuland 1989).

From monocausality to multiple causation

Three eras in the recent history of medical causality

In a rather simplified way, causation involves the relationship between at least two entities, an agent and a disease. Both can be easy to define and identify, or, on the opposite, they can be “fuzzy sets”, i.e., have blurred bounds. From this point of view, we can describe three eras in the history of medical causality in the last two centuries. The first era corresponds to the microbiological revolution, i.e., the triumph of a linear mono-causal (Aristotelian) concept of cause. After the work of Pasteur and Koch, the agent of a disease was conceived as a single *necessary* cause (e.g., *Mycobacterium* for tuberculosis). The concept of necessary cause means that the disease does not develop in the absence of exposure to the agent. Such a view implies: (a) that the cause is defined univocally and is easily identifiable; (b) that the disease can be also defined univocally, i.e., it is not a complex and variable constellation of symptoms. Sometimes such conditions clearly occur and the relationship between a (necessary) cause and the corresponding disease is evident: for example, smallpox is a clear-cut disease entity, easy to define and di-

agnose; it is due to a single necessary virus (no smallpox develops in the absence of the specific virus); and clear proof of the causal link has come from the disappearance of smallpox after large scale vaccination. On other occasions, the disease itself is at least partly defined on the basis of its cause: for example, from the symptomatologic point of view tuberculosis is a complex constellation, and the only unifying element has been the ability to identify *Mycobacterium* directly (microscopically in the lesions) or indirectly (immunologically). In the case of smallpox the symptomatology of the disease is so characteristic, and specificity of the relationship between cause and effect is so high that referring to the poxvirus as a necessary cause seems natural. In the case of tuberculosis, instead, the necessary character of the cause is weaker, since it comes from grouping part of the disease manifestations on the basis of the cause.

Situations like smallpox are a minority; more frequently, in the “Pasteur-Koch” paradigm we find a clearly defined agent (usually a bacterium, a parasite or a virus) which is used as the “unifying element” of a constellation of symptoms, i.e., the disease itself is largely defined and recognised on the basis of the agent. The popularity of the “Pasteur-Koch” approach to causality has not decreased, and the concept of a necessary cause of disease has been proposed still very recently as a universal paradigm in medicine (Sutter 1996).

The second era refers to chronic affections like cancer or cardiovascular disease. In this case the concept of “necessary” is not applicable on the basis of current knowledge. No “necessary” cause of cancer is known; rather, the idea of a “causal web” has been introduced and largely applied. The idea of a causal web implies that to induce the disease, the concurrence of different “exposures” or conditions is required, none of which is necessary. For example, lung cancer can be induced by a causal web including tobacco smoking and individual predisposition based on the CYP1A1 genotype. Another causal web may be represented by asbestos exposure and low consumption of raw fruits and vegetables in the occurrence of mesothelioma. The idea of the web implies that, while the disease is usually well defined from a clinical point of view (e.g., lung cancer or mesothelioma), causal agents are classified according to a “polythetic” classification: cases of lung cancer are not characterised all by the same exposure, but they share partially overlapping constellations of causes. The idea of “polythetic” classification corresponds to Wittgenstein’s definition of “a long rope twisted together out of many shorter fibres”.

The concept of causal web was already introduced in philosophy of causality by the British philosopher John Mackie, who coined the definition of INUS (Insufficient non-redundant

component of unnecessary sufficient complex) (Mackie 1965). For example, in explaining why a fire burned down the house, one can identify several components of a causal complex, like the strong wind, the fact that the electric oven was turned on, and the fact that the alarm system did not work. In such a complex, which was sufficient to start the fire, at least one component is non-redundant (i.e., it is an INUS condition), for example the failure of the alarm system (on the contrary, the strong wind is not an INUS). The definition of INUS corresponds to the logic of “*conditional counterfactuals*”, i.e., to asking for each component of the causal complex whether in its absence the effect would have developed anyway. Clearly, the idea of INUS still corresponds to the conception of causes as *necessary* events, although in the context of causal complexes.

In chronic disease epidemiology we have causal complexes without single necessary components. However, this is true only if we consider causality at the *individual level*: it is impossible to identify the necessary cause that explains the occurrence of a single case of cancer, while it is possible to identify a non-redundant component in the causal web that has led to the fire. If we shift from the individual to the *population*, then the idea of “non-redundant” component makes sense. If we consider this century’s epidemic of lung cancer, there is no doubt that it is attributable to the diffusion of the habit of smoking: although we cannot attribute each single case of lung cancer to the individual’s smoking habits, we are sure that on a population level the epidemic would not have occurred without cigarette smoking (*conditional counterfactual*). The risk of cancer in those who stop smoking decreases considerably in comparison with continuing smokers, and reaches after a few years the risk of non-smokers. Apparently, therefore, we have to apply different criteria of causation if we consider the individual level or the population level. We can say that for chronic diseases the INUS model is valid at the population level.

The third era is even more complicated than that. In the case of diseases like schizophrenia, bulimia or anorexia, both disease and agent have blurred bounds. The disease cannot be easily distinguished from other similar symptomatological constellations (for example, bulimia from other conditions characterised by obesity and “binge eating”), and the causal complexes are rather ill-defined and vague. In this case we shift from the classical scientific paradigm of “*explanation*” to a more evasive and slippery paradigm of “*understanding*” (Von Wright 1957) as used by psycho-social sciences. What should be clear is that there is a continuum between the three categories in Tab. 1, and diseases like smallpox are only one extreme of the spectrum, the other extreme being represented for example by several psychic disorders.

The importance of interactions: Helicobacter Pylori and gastric lymphomas

The example I wish to describe in this paragraph is relevant from several points of view. First of all, it refers to a causal relationship in which both exposure and disease are “polythetic” (group 4 in Tab. 1). In fact, H. pylori is likely to be causally associated with gastric lymphomas, but it is not the only risk factor and it is certainly not a sufficient cause. Secondly, it is an example of the complex role an infectious agent can play (far from being the necessary and sufficient cause of a specific disease). Third, the example illustrates both the difficulties of identifying risk factors by simple geographic correlations, and the importance of “*modifying factors*” (or susceptibility factors) in carcinogenesis.

Non-Hodgkin’s lymphomas (to which gastric lymphomas belong) are an extremely heterogeneous category of affections: some of them have a mild clinical course, others very rapid; histologically, they show a wide range of manifestations, and even their cells of origin are disparate (B- or T-lymphocytes). In fact there is not a simple, univocal and agreed upon classification: the recent REAL classification seems to be an operational tool rather than a new interpretation based on scientific evidence. Non-Hodgkin’s lymphomas (including gastric lymphomas) are on the rise in all the Western world at a rate of 3–4% per year, but the reasons for such an increase are unknown. Better and earlier diagnosis is not an explanation.

As far as the alleged “*cause*” of gastric lymphomas (H. pylori) is concerned, things are no simpler. There are several different antigenic varieties of H. pylori, apparently with different biological activities. A disease which is certainly due to H. pylori is peptic ulcer. The relationship between H. Pylori and lymphoma would *not* be demonstrated on the basis of “*geographic pathology*”. In fact, H. pylori infection is extremely frequent (it affects around 50% of the population), and shows wide geographic variability, while non-Hodgkin’s lymphomas are rare and show more limited geographic

Table 1 Diseases can be classified according to the nature of causal agents and to the appearance of signs/symptoms: in both cases monothetic or polythetic definitions are possible

	Disease	
	Monothetic definition ^a	Polythetic definition ^a
Agent (s)		
Monothetic definition	Smallpox (group 1)	Tuberculosis (group 2)
Polythetic definition	lung cancer (group 3)	bulimia and anorexia (group 4)

^a On the basis of clinical signs and symptoms only.

variation. There is very little overlapping of the distributions of these two conditions: 85% of the Indian population, for example, is infected with *H. pylori*, but the rate of lymphomas, including those of the stomach, is lower than in Western populations. In the United States, the prevalence of *H. pylori* infection and the mortality rate for peptic ulcer have steadily decreased during the past 50 years, but Non-Hodgkin's lymphomas are increasing. So, on the basis of geographic or time correlations, one would *not* conclude that *H. pylori* is a cause of gastric lymphomas. However, we have much more stringent data. Epidemiological prospective studies from distinct populations have reported that the risk of gastric lymphoma in patients with antibodies against *H. pylori* is about six times higher than in normal subjects, while there is no association with non-gastric lymphomas. More interestingly, eradication of *H. pylori* with antibiotics lead to regression of low-grade gastric lymphomas; the latter is a kind of experimental "galileian" evidence which strongly supports the causal hypothesis. Incidentally, even in the absence of such experimental evidence it would not be easy to argue against epidemiological proof. To explain away a relative risk of six, in fact, one has to suppose that an alternative exposure exists, which has a *stronger* association with both the disease and *H. pylori*: it is possible, but very unlikely (IARC 1994).

Therefore, one has to admit that a very common infection is responsible for a very rare disease. This is equivalent to admitting that something else interacts with *H. pylori* in order to explain its causal role. This idea of interaction is not new. Table 2 shows that also in the case of a disease belonging to group 2 in Table 1 (tuberculosis), host response is crucial: the risk of developing TB is in fact much higher in strict relatives of cases. In addition to genetic susceptibility, there are several other types of "acquired" interactions. For example, interaction occurs between cigarette smoke and asbestos in modifying the risk of lung cancer (Tomatis 1990).

Table 2 Effect of genetic relatedness on host response to *M. tuberculosis* in families with an index case (from Evans 1993)

Relation of family member to index case	% of exposed and susceptibles showing clinical manifestations of TB
Marriage partner	7.1
Half-sibling	11.9
Dyzygotic twin	25.5
Monozygotic twin	83.3

The biological background of epidemiological observations

The DNA dogma in carcinogenesis

I already anticipated that in the current interpretation of carcinogenesis and of cause-effect relationships in cancer epidemiology, inferences are made against a background represented by the idea of a "homeostatic" or feedback mechanism involving genotoxic "hits" from the environment, on one side, and DNA repair on the other side. Discussions on causality are not in a vacuum, but are strictly conditioned on the current models of disease.

Centrality of DNA in modern biology has reached the status of a "dogma". Contemporary, DNA-centred molecular biology was born from a strict interaction with physics. Seminal work was represented in 1994 by Erwin Schroedinger's book "What is life?" (Schroedinger 1994), where he suggested very clearly that life phenomena could be explained by the properties of molecules, i.e., by their "memory". Schroedinger, a well-known quantum physicist with biological interests, was responsible for creating a bridge between the two disciplines and for the role that was attributed to DNA subsequently. In addition, he had views that were very similar to those already expressed (one century before) by Claude Bernard, updated with a reference to the theory of information: the mechanism by which an organism is kept stable consists, according to Schroedinger, in "absorbing order from the environment". In other words, Schroedinger interpreted homeostasis mainly as an exchange of information between internal and external molecules.

As far as the central role of DNA in contemporary biology is concerned, the relationship between "cause" and "mechanisms" sometimes is misunderstood. Molecular biologists tend to consider relevant changes at the molecular level as the "cause" of cancer, while epidemiologists usually refer to external agents as genuine causes. Perhaps it is not irrelevant to refer to Aristotle's four categories of cause: *material* (the cells in which cancer arises), *final* (i.e., the scope of malignant transformation, which can be described – although improperly – as a selective advantage of cancer cells, conferring them the ability to overcome the host's defences), *formal* (the morphological characteristics of cancer and the corresponding functional changes), and *efficient* (the events that trigger the mechanistic steps which lead to malignancy). Molecular changes at the DNA level refer to the "formal" cause, the molecular and functional changes that occur in a normal cell, while external agents refer to the "efficient" cause.

In spite of such distinctions, similar criteria for "causality" assessment can be applied to both external causes and

mechanisms. The so called “Henle-Koch postulates” have been used for a long time to describe causality assessment in medicine. They were derived from the discoveries of the microbiological era and stated that an agent is a cause of disease if it is present in all the affected persons (“necessary” cause), it is absent in healthy subjects (“sufficient” cause) and can be inoculated into an animal to induce the same disease that it causes in humans. In fact, the long-used definition of “Henle-Koch postulates” is wrong, both because they are not postulates, and because they were changed considerably between the original formulation by Henle and the one due to Koch.

Recent developments in molecular biology seem to fit rather well with Henle-Koch’s rules, for example in the case of oncogenes. Oncogenes are mutated genes that, in their normal (wild) form exert crucial functions in the cell’s metabolism, and are highly conserved on the evolutionary scale. Mutated oncogenes have been found in a high proportion of malignancies and some of them have marked similarities with viral DNA sequences that are able to induce malignancies. Though the criterion of “necessary” cause does not seem to be met (since not all malignancies show oncogene mutations), this could be attributed to limited knowledge of relevant genes and limited sensitivity of tests. Both the criterion of “sufficient” cause, and the criterion of experimental reproducibility (at least in particular circumstances) are met, in that “transfection” of NIH 3T3 cells (which already underwent partial transformation) with a mutated oncogene confers malignant properties.

The nature of medical theories and conditionalized realism

What is the nature of an observational medical theory, such as “tobacco smoking causes lung cancer”? It can hardly be claimed that such theories represent universal laws of nature, comparable to the laws of thermodynamics or molecular genetics. In the meantime, they cannot be dismissed as simple empirical generalisations. We believe that the statement “smoking causes lung cancer” is clearly related to some natural phenomenon. The feeling that it reflects something more than an empirical generalisation does not mean that we are ready to accept that such statement is comparable to laws describing basic natural phenomena like the genetic code.

According to Schaffner (1993), biology is characterised by “middle range” theories, i.e., laws that are intermediate between the simple observation of empirical regularities and universal statements about nature. Such middle range theories have the peculiarity of being strongly based on mutual reinforcement between different types of evidence, at different levels of reality and including some reference to basic

laws of nature. The two main features of middle range theories are their being *temporal* models (i.e., they refer to phenomena that undergo a process, like carcinogenesis) and their being “overlapping inter-level models” (i.e., they serve to connect different levels of reality).

Let us consider the relationship between tobacco and cancer (Vineis & Caporaso 1995). Even after the publication of persuasive evidence linking lung cancer to tobacco smoking, some investigators questioned whether the epidemiologic evidence incriminated smoking as a cause of cancer in humans. In particular, R.A. Fisher, an eminent statistician of this century, claimed that the early epidemiological observations could not be interpreted as a proof of cause-effect relationship, arguing that one could not rule out that a genetic factor both increased the propensity to smoke and the risk of lung cancer. A key criticism was that exact knowledge of the mechanisms of tobacco carcinogenesis was necessary to establish a cause-effect relationship. Such criticism was at the root of scepticism towards epidemiological evidence and its applications in public health.

In fact, in addition to the (redundant) epidemiological observations linking tobacco to lung cancer in humans, we have several types of evidence at different levels. Tobacco smoke contains many mutagenic and carcinogenic substances. Both tobacco smoke and extracts induced tumours in experimental animals. A general trend in molecular studies is the increasing evidence that point mutations in tumour suppresser genes (i.e., p. 53) and oncogenes (i.e., *ras*) may be specific both for the type of tumour and for the critical environmental exposure; this is true also for tobacco.

Furthermore, Fisher’s hypothesis that genetic predisposition both induces smoking habits and increases the risk of lung cancer has been refuted on the basis of twin studies (Vineis & Caporaso 1995).

To admit that smoking causes lung cancer one need not be either a *realist* or an *empiricist*, to refer to a long-lasting debate in medicine. The realist postulates that empirical observations do refer to some reality in the external world (independently of theoretical models); the empiricist strictly sticks to observable entities, avoiding any judgement about the essence of reality. For example, realists in medicine tend to believe that basic biochemical or molecular mechanisms explain the effectiveness of therapies; while empiricists strongly advocate empirical evidence coming from randomised controlled trials. Wide areas of observational medicine, and particularly epidemiology, clearly belong to the empiricist field. As a third alternative, Schaffner (1993) proposes a “*conditionalized realism*”. This means that a “middle range” theory is held to be true if two conditions are met: (1) that also “auxiliary hypotheses” are true; (2) that no valid

alternative explanation can be put forward. The second condition is well known to epidemiologists, since it corresponds to the concept of “confounding”. The first condition is also easily understandable: examples of auxiliary hypotheses are that the design of a particular study did not introduce bias; that the evidence collected from animal experiments can be extrapolated to humans; that tobacco-related mutations in specific genes (oncogenes) actually are relevant to the carcinogenic process.

Which type of message does the “tobacco and cancer” example convey? *First*, we believe that smoking causes cancer not only on the basis of empirical observations in humans (which are limited by their non-experimental nature), but also because we have independent proof referring to different levels of reality. Such proof includes reference to some of our most profound beliefs concerning nature, such as the crucial role played by DNA damage in carcinogenesis. Therefore, prior beliefs in nature are essential in the interpretation of empirical observations.

Secondly, the model of causality which is valid in observational medicine is compatible with the models that have been proposed for physics (such as the INUS model), particularly if we refer not to the individual (human or molecule), but to populations. *Third*, as in other fields of science, also in observational medicine the truth of a theory is conditionalised on auxiliary hypotheses and the lack of alternative explanations. This conditionalised nature of biologic realism (Schaffner 1993) is an example of the interplay between direct evidence and interpretation, in that even an experiment – such as a randomised trial – will be interpretable only in the context of background knowledge concerning auxiliary hypotheses (although a randomised experimental trial needs *less* auxiliary hypotheses than observational medicine).

How strictly interpretation and prior belief can be intertwined with the scientific practice of observational medicine is shown by the role of “model selection” in causal inference. According to Robins and Greenland (1986), “all modelling strategies contain implicit prior beliefs about nature”. When we choose which “explanatory” variables, confounders or effect modifiers to include into the occurrence function, we anticipate which of them make biological sense and are compatible with a reasonable interpretation of the data. “[...] a statistical model is a mathematical expression for a set of assumed restrictions on the possible states of nature [...]. For example, a linear [...] logistic model for the dependence of

subsequent fertility on dibromochloropropane exposure and parity implies the following restrictions about nature: (1) an exponential dependence of the fertility odds ratio on DBCP and parity; (2) a constant odds ratio across DBCP for the association of any parity level [...] with subsequent fertility; and (3) a constant odds ratio across parity for the association of any DBCP level with subsequent fertility” (Robins & Greenland 1986). The choice of the model, in fact, is a trade-off between different and potentially conflicting goals, such as “saving variance” in the model by introducing *few* assumptions about nature, decreasing bias by introducing *correct* assumptions, and increasing bias by introducing *incorrect* assumptions. In fact, an occurrence function based on a “highly-saturated” model will have small bias provided that no confounding remains, but may have large statistical variance.

Therefore, any opposition between scientific knowledge (based on the observation of facts within an “occurrence function”), and non-scientific prior beliefs would be misleading, since prior belief is clearly necessary for a correct building and interpretation of causal models.

Conclusions

Knowledge of the causes of a disease allows therapeutic and preventive interventions in humans. Due to such practical implications (the well-being of the patient or the population), medicine is not entirely a natural science. Rather, both medicine and epidemiology are at the cross-roads of natural sciences and human sciences.

Epidemiology has evolved from a mono-causal to a multi-causal concept of the “web of causation”, thus mimicking a similar and much earlier shift in the social sciences. However, in comparison with the social sciences epidemiology is both more sensitive to *underlying biological models* (which condition the interpretation of population findings), and more prone to a *simplification* of the causal pathways. For example, epidemiology has developed more sophisticated theoretical models for bias and confounding than the social sciences did, thus revealing the practical purpose of identifying single preventable risk factors. Epidemiology makes use more often of study designs that simulate experimentation (see also “Bias”), than of surveys in the general population. Epidemiology, therefore, represents an interesting and unique example of cross-fertilisation between social and natural sciences.

Zusammenfassung

Kausalität in der Epidemiologie

Die Epidemiologie ist ein interessantes und einmaliges Beispiel einer „Kreuzbefruchtung“ von Sozial- und Naturwissenschaften. Die Epidemiologie hat sich von einem monokausalen zu einem multikausalen Konzept eines „Ursachennetzwerkes“ entwickelt. Genau dieselbe Entwicklung vollzog sich in den Sozialwissenschaften, aber viel früher. Im Vergleich mit den Sozialwissenschaften, geht die Epidemiologie aber einerseits vorsichtiger mit zugrunde liegenden biologischen Modellen um (welche die Interpretation von Bevölkerungsdaten bedingen) und neigt andererseits dazu, ursächliche Zusammenhänge zu vereinfachen. So hat die Epidemiologie zum Beispiel viel differenziertere theoretische Modelle für Bias (Verzerrung) und Confounding (Störfaktoren) entwickelt als dies in den Sozialwissenschaften der Fall ist, was mit der praktischen Zielsetzung zusammenhängt, einzelne vermeidbare Risikofaktoren zu identifizieren. Die Epidemiologie macht häufiger Gebrauch von Studiendesigns, die Experimente simulieren, als von Befragungen in der Allgemeinbevölkerung.

Résumé

Causalité en épidémiologie

L'épidémiologie représente un exemple intéressant et unique de fertilisation croisée entre les sciences sociales et les sciences naturelles. L'épidémiologie a évolué d'un concept monocausal au concept multicausal de la "toile de causalité", copiant ainsi une évolution similaire mais plus précoce dans les sciences sociales. Cependant, par rapport aux sciences sociales, l'épidémiologie est plus sensible aux modèles biologiques sous-jacents (qui conditionnent l'interprétation des observations populationnelles) et est aussi plus portée à la simplification des circuits causaux. Paradoxalement, l'épidémiologie a développé des modèles théoriques plus sophistiqués pour décrire les biais et les effets de confusion que les sciences sociales ne l'ont fait, mais dont la raison pratique était d'identifier des facteurs de risque uniques et modifiables. L'épidémiologie recourt plus souvent aux plans d'étude simulant l'expérimentation qu'aux enquêtes dans la population générale.

References

Bernard C (1957). Introduction à l'étude de la médecine expérimentale (1865). Transl. Green H.G. New York: Dover.

Evans AS (1993). Causation and disease: a chronological journey. New York: Plenum Medical Book Co.

Hempel CG (1966). Philosophy of natural science. Englewood Cliffs, NJ: Prentice-Hall.

IARC, International Agency for Research on Cancer (1994). IARC Monographs on the evaluation of carcinogenic risks to humans. Lyon: IARC. (Vol. 61).

Mackie JL (1965). Causes and conditions. Am Philosoph Quart 2: 245–55.

Miettinen OS (1985). Theoretical epidemiology. New York: John Wiley and Sons.

Nuland SB (1989). Doctors: the biography of medicine. New York: Vintage Books.

Odlroyd D (1986). The arch of knowledge. New York: Methuen.

Robins J, Greenland S (1986). The role of model selection in causal inference from nonexperimental data. Am J Epidemiol 123: 392–402.

Schaffner KF (1993). Discovery and explanation in biology and medicine. Chicago: Chicago University Press.

Schroedinger E (1994). What is life? Cambridge: Cambridge University Press.

Sutter MC (1996). Assigning causation of disease: beyond Koch's postulates. Persp Biol Med 39: 581–92.

Tomatis L, ed. (1990). Cancer: causes, occurrence and control. Lyon: IARC (IARC Scie. Publ.; no. 100).

Vineis P, Caporaso N (1995). Tobacco and cancer: epidemiology and the laboratory. Environ Health Persp 103: 156–60.

von Engelhardt D (1993). Causality and conditionality in medicine around 1900. In: Delkeskamp-Hayes C, Gardell Cutter MA, eds. Science, technology, and the art of medicine. Dordrecht: Kluwer: 75–104.

Address for correspondence

Prof. Paolo Vineis
Unit of Clinical Epidemiology
Ospedale S. Giovanni Battista
via Santena 7
I-10126 Torino
Phone: +39-011-670-6525
Fax: +39-011-670-6692

e-mail: paolo.vineis@unito.it