History and Philosophy of Modern Epidemiology

Based on a talk delivered at the &HPS Conference, Pittsburgh, October 2007

Hanne Andersen
Department of Science Studies, University of Aarhus
DK-8000 Århus C, Denmark

Abstract: Epidemiological studies of chronic diseases began around the mid-20th century. Contrary to the infectious disease epidemiology which had prevailed at the beginning of the 20th century and which had focused on single agents causing individual diseases, the chronic disease epidemiology which emerged at the end of Word War II was a much more complex enterprise that investigated a multiplicity of risk factors for each disease. Involved in the development of chronic disease epidemiology were therefore fundamental discussions on the notion of causality, especially the question when causal inferences could be justified. In this paper, I shall analyze the implicit normativity of these debates.

First, I shall give a brief overview of the historical background on which chronic disease epidemiology emerged and describe how the pioneer studies on smoking and lung cancer became icon of the major challenge that the emerging chronic disease epidemiology was facing: the impossibility of proving that statistical associations reflected causal relations. Next, I shall describe how the development from the monocausal enterprise of infectious disease epidemiology to the multicausal enterprise of chronic disease epidemiology gave rise to intense discussions of the possible criteria by which to establish causal relationships between a given factor and a particular disease. I shall show how the necessary and sufficient conditions expressed in the so-called Henle-Koch criteria that had proved useful for the 19th century investigations of infectious diseases remained an ideal, although clearly an unobtainable one. Thus, I shall show how 20th century chronic disease epidemiologists on the one hand were searching for a new set of general principles which would provide a logical framework for their investigations, but on the other hand admitted that they would have to accept something more "pragmatic". I shall analyze the various positions in this debate, arguing that the implacability of the debate was due to unrecognized normative issues. I shall argue that many insisted on a distinction between science and application that was untenable, but that due to this distinction the values involved in deciding whether or not to act on the basis of a hypothesis were rarely explicitly discussed and the decision therefore continued to appear as a matter of taste rather than the result of a cogent normative analysis.

1. The rise of chronic disease epidemiology

Chronic disease epidemiology emerged as a new research field in the middle of the 20th century. During the latter half of the 19th and the first half of the 20th century, significant changes in the morbidity and mortality patterns had taken place in most

Western countries. Due to such developments as the creation of filtered water supplies and sewage systems as well as the public health movement's emphasis on cleanliness, serious epidemic diseases like typhus and tuberculosis had retreated steadily during the 19th century. Discoveries of vaccines for diseases like diphtheria and typhoid fever by the late 19th century or tetanus around World War I contributed further to the decrease in mortality rates, as did the discovery of sulpha drugs and the discovery and development of mass production methods of penicillin during World War II. But above all, living standards and nutrition had increased considerably and contributed to the steady decline of infectious diseases during the first half of the 20th century.

As a result of all these developments, by the mid-20th century many of the worst infectious diseases had come more or less under control. The sickness that affected society was therefore no longer primarily produced by infections, but rather by diseases such as cancer, coronary artery disease, diabetes, ulcers, and strokes. It was these "chronic" diseases that now represented an increasing threat to the population, especially the middle-aged men among whom the incidence of e.g. ischemic heart disease was growing by such a rate that it became described as a "modern epidemic" (Morris 1964). Chronic diseases therefore became a new target of attention for epidemiologists.

During the late 19th and early 20th century, the infectious disease epidemiology had focussed on single agents causing individual diseases. In contrast, the chronic disease epidemiology which was now called for was a much more complex enterprise that would have to investigate a multiplicity of risk factors for each disease. This shift from the mono-causal enterprise of infectious disease epidemiology to the multi-causal enterprise of chronic disease epidemiology was linked to several methodological developments. These including refinements of fundamental concepts like bias and confounding, and the development of advanced study designs such as case-control and cohort designs. Admittedly, the awareness of fallacies related to e.g. self-selection or inter-mixed causal effects has a much longer history than the chronic dis-

ease epidemiology, but the notion of confounding and the distinction between various forms of bias were developed and refined as an important part of the maturing chronic disease epidemiology (Vandenbroucke 2002; Vineis 2002). Likewise, although case-control and cohort studies had a history that went back at least to the 19th century, these designs underwent rapid development and refinement in the decades after World War II. During this period of methodological development and refinement, the Framingham Study on cardiovascular disease (Dawber 1980; Gordon and Kannel 1970), the cohort study of British doctors (Doll and Hill 1956) and the case-control study of cancer cases from London hospitals (Doll and Hill 1950a) came to be seen as exemplars of the methodology of the new chronic disease epidemiology (Paneth, Susser, and Susser 2002a; Paneth, Susser, and Susser 2002b; Susser 1985).[1]

2. The first major studies: smoking and lung cancer

Lung cancer was one of the first objects of investigation. Lung cancer rates had been seen to increase dramatically in many parts of the world in the 1930es (Doll and Hill 1950a), and there were several hypotheses on the causes for this increase, including car exhaust fumes, gas works, industrial pollutants, coal fires, road tars, delayed reaction to the Spanish flu, and cigarette smoking. During the 1940s and 1950s, several studies were initiated to examine the possible link between smoking and lung cancer. The first studies were retrospective case-control studies that used hospital populations (Wynder and Graham 1950; Levin, Goldstein, and Gerhardt 1950; Doll and Hill 1950b), but prospective studies were soon to follow (Hammond and Horn 1958a; Hammond and Horn 1958b; Hill and Doll 1956; Doll and Hill 1956). The studies

[1] For actors' accounts of the development of modern epidemiology, see e.g. (Susser 1985; Susser 1996b; Susser 1996a). As Berlivet (2005) has noted there is a tendency in the historical accounts of the development of epidemiology to describe a 'rupture' by the middle of the 20th century which overemphasized the transformation of the discipline and downplays the importance of the historical predecessors. However, the focus of this paper is not on the relation to the historical predecessors and the question of continuity versus discontinuity, but solely on how causality was discussed among the participants in the early development of chronic disease epidemiology.

showed a clear association between smoking and lung cancer, and during the 1950es the view gradually developed that smoking was causing lung cancer.[2]

However, objections were raised by a number of prominent biostatisticians and epidemiologists,[3] most notably the British biostatistician Ronald Fisher (Fisher 1957a; Fisher 1958c; Fisher 1958a; Fisher 1959; Fisher 1958b) and the American biostatistician Joseph Berkson (Berkson 1955; Berkson 1958). Fisher had been one of the main architects of the randomized clinical trial (RCT) that had been introduced to medicine during the 1940s and 1950es. Gradually, the RCT had gained a role as the 'gold standard' of medical research together with the idea that bias and confounding could be avoided by randomization (Vineis 2002). Thus, in his seminal book *The Design of Experiments* (Fisher 1935) Fisher had argued that since it was not possible to eliminate all sources of variation other than those under test, randomization was the essential safeguard against corruption of the results due to uncontrolled differences. In a similar vein, Fisher argued against a causal interpretation of the observed association between smoking and lung cancer on the grounds that lung cancer and smoking might both be produced by a common factor, such as, for example, a common genetic trait (Fisher 1957a; Fisher 1958c).[4] Likewise, Berkson, who had been one of the key scholars in developing the understanding of confounding (Berkson 1946), warned that the observed associations could be spurious (Berkson 1958), and like Fisher he also stressed the possibility of a common constitutional cause of the association.[5]

---

[2] On the history of the smoking and lung cancer research, see e.g. Lock, Reynolds, and Tansey 1998; Talley, Kuschner, and Sterk 2004; Kluger 1996; Brandt 1990; Doll 1998.

[3] An extensive literature exists on this debate, both from historical actors, e.g. Doll 2002; Doll 2004; Wynder 1997 and from later epidemiologists and historians, e.g. Hill, Millar, and Connelly 2003; Stolley 1991; Parascandola 2004a; Parascandola 2004b.

[4] Fisher even suggested the interpretation of the association between smoking and lung cancer that lung cancer could be the cause of cigarette smoking; a suggestion that was described by his peers as "advanced .. apparently for the sake of logical completeness, and it is not clear whether it is intended to be regarded as a serious possibility" (Cornfield et al. 1959).

[5] Parts of Berkson's argument also seemed to be advanced more out of logical possibility than actual plausibility; something that he was aware of and overtly defended with reference to the general lack of knowledge of the disease mechanisms that ought to leave openness for suggestions in all directions:

Although Fisher and Berkson have become notorious for their strong opposition to the hypothesis of a causal relation between smoking and lung cancer and for their strong rhetoric in defence of their position, their concerns were shared by several other biostatisticians. As argued by Parascandola (Parascandola 2004a; Parascandola 2004b), part of the debate was aimed at defending the authority of the biostatistician in medical science by emphasizing both the importance of a rigorous empirical methodology and the dangers of the inferential reasoning employed by the epidemiologists. In a lecture Fisher stressed that "I should be the last person to attack evidence for being merely statistical, because for a great part of my work I have been concerned with the problem of *how* experimentation should be carried out, *how* reasoning processes should be applied to the data supplied by experimentation or by survey so as to really give conclusive answers" (Fisher 1958b). On his view, the key to achieving such "entirely unchallengeable conclusions" (ibid. p. 12) were controlled, randomized experiments.

But apart from the objections related to confounding and bias, two other objections were frequently seen, namely first, the fact that smoking seemed to be associated not only with lung cancer, but with a multitude of other diseases as well, and second, the lack of a proven carcinogenic agent. The fact that smoking was associated not only with lung cancer, but with a number of other diseases, including coronary artery disease, pulmonary tuberculosis, and peptic ulcer was raised as a cause of suspicion against a causal relation. For example, Berkson argued that

> I find it quite incredible that smoking should cause all these diseases. It appears to me that some other explanation must be formulated for the multiple statistical associations found with so wide

---

"After all, the small group of persons who successfully resist the incessantly applied blandishments and reflex conditioning of the cigaret advertisers are a hardy lot, and, if they can withstand these assaults, they should have relatively little difficulty in fending off tuberculosis or even cancer! If it seems difficult to visualize how such a constitutional influence can carry over to manifest itself as a graded increase of death rate with a graded increase of intensity of smoking, then we must remember that we are wandering in a wilderness of unknowns. I do not profess to be able to track out the implications of

> a variety of categories of disease. And if we are not crassly to violate the principle of Occam's razor, we should not attribute to each separate association a radically different explanation (Berkson 1958).

Others argued that since smoking seemed to be a cause of all ills, it was more likely to be a symptom rather than a cause, and that it was possible that "the type of person who is careful with his health is less likely to be a cigarette smoker and that the cigarette smoker is likely to be the person who generally takes greater health risks" (Arkin 1955; quoted from Cornfield et al. 1959). However, these arguments were only formulated loosely as implausibility arguments. Proponents of a causal relation countered by recollecting similar cases such as the Great Fog of London that had also caused a multitude of diseases:

> We see nothing inherently contradictory nor inconsistent in the suggestion that one agent can be responsible for more than one disease, nor are we lacking precedents. The Great Fog of London in 1952 increased the death rate for a number of causes, particularly respiratory and coronary disease, but no one has given this as a reason for doubting the causal role of the Fog. Tobacco smoke, too, is a complex substance and consists of many different combustion products. It would be more 'incredible' to find that these hundreds of chemical products all had the same effect than to find the contrary. A universe in which cause and effect always have a one-to-one correspondence with each other would be easier to understand, but it obviously is not the kind we inhabit (Cornfield et al. 1959).

By the same token, it was repeatedly emphasized that the difference between the specific relations of infectious agents to individual diseases, and the apparent non-specific relations of agents like tobacco to a multiplicity of diseases primarily reflected only "differences in levels of development of knowledge of these two types of disease, rather than differences in the logic of the situation" (Lilienfeld 1959). For example, a vector like polluted water would be associated with several infectious dis-

---

the constitutional theory or to defend it, but it cannot be disposed of merely by flat denial" (Berkson

eases, and if seen as a vector carrying a multitude of chemicals, tobacco was no different.

However, whereas the specificity objection could be countered by examples and simple plausibility arguments, the last argument, the lack of a carcinogenic agent that could be shown in the laboratory to affect cells to become cancerous, was by many seen as much more severe. Doll and Hill had noted in their initial report that "as to the nature of the carcinogen we have no evidence" (Doll and Hill 1950a). Early attempts to induce cancer in animals with tobacco products had been unsuccessful, and although new experiments performed in the early 1950es did indicate that cancer could be induced by tobacco tar (Wynder, Graham, and Croninger 1953), reservation remained. Thus, in an editorial in BMJ it was noted that "experimental work has not provided complete and irrefutable proof" and that this had "tended to hinder ... whole-hearted acceptance" of the smoking-lung cancer hypotheses (anon. 1958).[6]

Similar to Fisher and Berkson's concerns that the observational studies could not rule out confounding or bias, the objection directed at the multiplicity of diseases associated with smoking as well as the concern over the lack of a proven carcinogenic agent served not to refute the smoking and lung cancer hypothesis but simply to question the basis for making such a causal inference. Thus, the smoking and lung cancer debate remained unsettled, and in the Surgeon General's (Surgeon General 1964) report *Smoking and Health* it was clearly noted that although there was an association between smoking and lung cancer, there was no *proof* of a causal relation.

The lung cancer studies therefore became icon of the major challenge that the emerging chronic disease epidemiology was facing: the impossibility of proving that statistical associations reflected causal relations and, given this challenge, how to re-

---

1958).

[6] Similar concerns were expressed in the Lancet that reported of "a slight nagging uncertainty as to the evidence ... No classical carcinogen has been found in adequate concentration in tobacco smoke; no genuine lung cancers have been produced experimentally; and, though tobacco tar produces cancer when painted in mouse skin, it is a slow and ineffective agent by all ordinary standards" (anon. 1962). See also (Parascandola 1998) for details of this criticism.

concile scientific results encumbered with such uncertainties with the practical needs of health administrators to take preventive actions.

3. Coming to terms with multi-causality

The self-image of epidemiology had been formed by early infectious-disease epidemiologists who had managed to discover the disease transmitting vectors and saving lives by their removal. Legends like John Snow's removal of the handle of the Broad Street pump in London during the cholera epidemic in 1854 were at the core of epidemiologists disciplinary identity. One of the recurrent themes in the discussions of smoking and lung cancer was therefore how the study of chronic diseases in general compared to the much more simple studies of infectious diseases.

After the bacteriological revolution the Henle-Koch postulates had served as simple guidelines for establishing causal relations between infectious agents and disease. According to the Henle-Koch postulates, a given infectious agent could be said to be the cause of a specific disease if and only if 1) the agent occurred in every case of the disease and under circumstances which could account for the pathological changes and clinical course of the disease, 2) the agent occurred in no other disease as a fortuitous and non-pathogenic agent, and 3) after being fully isolated from the body and repeatedly grown in pure culture, the agent could induce the disease anew.[7] The Henle-Koch postulates thus established a classical regularity view of causality in which the two first postulates served to establish the infectious agent as a necessary and sufficient condition for the disease, while the third served to establish the direction of the relation.

Although it was clear that the Henle-Koch postulates expressed an idealized view, even for infectious diseases, the Humean regularity view of causality implicit

[7] The Henle-Koch criteria can be found in various, slightly different versions in the epidemiological literature. For an account of how the Henle-Koch postulates were handed down in the epidemiological literature, see e.g. (Evans 1976). On the many subtleties excluded from the traditional accounts, see e.g. (Carter 1985; Carter 1991). Criteria similar to the Henle-Koch postulates were also suggested by other bacteriologists, see e.g. (Summers 2007) for an account.

in them remained an ideal for many of the participants in the discussion. For example, two of the leading American biostatisticians, Jacob Yerushalmy and Carroll E. Palmer, phrased their methodological discussion of the investigation of chronic diseases as a direct parallel to the methodology of investigating infectious diseases:

> it may be instructive to compare the current approach in investigations of etiologic factors in many chronic diseases with the more rigorous methods long in use by bacteriologists in implicating a living organism as the causal agent for a specific disease. Almost from the very beginning, when bacteria were first found to cause disease, bacteriologists felt the need for a set of rules to act as guideposts in investigations of bacteria as possible causal agents in disease. The formulation by Koch of these postulates and their utilization in the field of bacteriology has contributed greatly to the orderly and systematic identification of causative organisms in many diseases. It is the purpose of this paper to develop an elementary parallelism between investigations of etiologic factors in certain chronic diseases and those of bacterial diseases (Yerushalmy and Palmer 1959).

This ideal was based on the view that further studies of the conditions apparently involved in the causation of a given disease – such as the condition of smoking involved in the causation of lung cancer – would show these conditions to be merely vectors containing what was truly the specific causative agent. Thus, once this agent could be found the demonstration that it caused the particular disease "would not differ essentially in methodology from the demonstration that a certain microorganism causes a specific bacterial infection" (Yerushalmy and Palmer 1959). On this view, multicausality was primarily a sign of the current ignorance.

Yerushalmy and Palmer were far from alone in adopting such a micro-reductionist view of disease causation. The American epidemiologist Abraham Lilienfeld, brother-in-law to Yerushalmy, described the same assumption that all disease causation would ultimately have to be described at a cellular level; a view that he saw as a "chain of causal relationships"

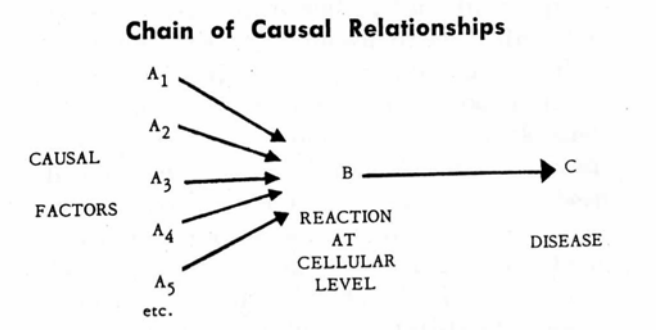

Figure 1 Chain of causation from Lilienfeld 1957

> To illustrate, the cause and effect relationships with multiple etiological factors, labelled A1, A2, A3, and so forth, each acting independently, are presented in the accompanying drawing. These factors can be looked on as producing a change in B at a cellular level. The changed cell B could then develop into C, the disease. Clearly, the cellular change in B can be considered as the necessary and sufficient condition for the disease C. Therefore, to meet the more rigorous definition of causality, the biological mechanisms relating A to B and B to C must be determined (Lilienfeld 1957).

On the one hand, this view of the 'chain of causation' did support the view that multicausality was in a way merely reflection of the chosen frame of reference, and that further investigations into the constituents of the vector would reveal a simple causal relation. But on the other hand, Lilienfeld also made clear that the idea of infectious agents as 'the specific causes' of the associated infectious diseases was no better than seeing vectors as causes. For example, considering the typhoid bacillus as the specific cause of typhoid fever or the dysentery bacillus as the specific cause of dysentery, Lilienfeld criticized Yerushalmy and Palmer's view of the bacilli as the definite causal agents claiming that "From a molecular viewpoint, the typhoid bacillus can also be considered a vector of a specific chemical agent which is the "real" cause of the disease" (Lilienfeld 1959, p. 42).

However, whether the 'specific cause' of a disease should be found on the level of bacteria and viruses or at the level of their biochemical constituents, the idea of a chain of causal relationships still implied a specific link in the chain that could be

seen as the necessary and sufficient condition for the further links in the chain. This idea of a *chain* of causal relationships was gradually questioned in favour of the much more complex idea of causal relations as a *web*. One of the major contributions in this development was 'the web of causation' introduced by MacMahon, Pugh and Ipsen in their classic textbook *Epidemiologic Methods* from 1960 where they argued against any distinction between different kinds of factors. From an example on the association between the treatment for syphilis and jaundice they depicted the many components that entered into the association as a complicated but still incomplete net, or as they termed it, "a web, which in its complexity and origins lies quite beyond our understanding" (MacMahon, Pugh, and Ipsen 1960)

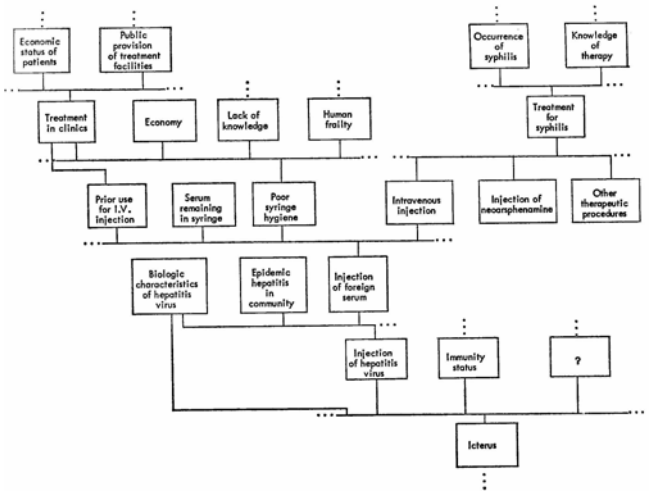

Figure 2 web of causation from MacMahon et al 1960

When the complexity and wide origin of this web was appreciated, it could be seen that, even for an infectious disease like syphilis, factors such as nutrition or genetics were as important as the infectious agent, and they suggested that

> the role of microbiologic agents in infectious disease has been considered predominant over that of genetic factors, perhaps in part because their importance became evident earlier but primarily because it is a good deal easier to control human water supply and to eradicate insect vectors than to breed genetically resistant populations (MacMahon, Pugh, and Ipsen 1960).

On this view, the precedence of one kind of factors, such as infectious agents or cellular mechanisms, reflected only human interests. Multi-causality was not simply a sign of ignorance that could be overcome through more research, but an inescapable condition.

4. Discussing guidelines

Given that chronic disease epidemiology could not establish any invariable relations of succession between causes and effects, it had instead to rely on a probabilistic conception of causality. This meant, first, that the suspected characteristic should be found more frequently in persons with the disease than in persons without the disease in question, and second, that persons possessing the characteristic should develop the disease more frequently than persons not possessing the characteristic (cf. Yerushalmy & Palmer p. 32, similar Lilienfeld 1957, p. 51).

However, like for necessary and sufficient conditions, this simple probability raising requirement entailed only that two characteristics were associated, not that the one was the cause of the other, or that they were not both the effect of some third factor. For infectious diseases, the third of the Henle-Koch postulates, the requirement of an experimental induction of the disease from pure culture, would take care of these two points. In accordance with the experimental and manipulative element of this postulate, it was often noted that the only certain way of establishing an association between a disease and a characteristic would be properly designed human experiments. This ideal of experiments was supported not only by the recurrent recollections of the triumphs of bacteriology, but also by the recent triumphant progress of the controlled trial which had been disseminated by some of the major epidemiologists, such as the British biostatistician and epidemiologist Bradford Hill (e.g. Hill 1951; Hill 1952; Hill 1953). Further, although controlled trials had primarily been associated with therapeutics, recent studies, like the Grand Rapids study of fluorine in

drinking water (Dean et al. 1950; Dean 1992), had shown how controlled trials also had a role to play examining preventive measures through the kind of community studies that was part of the core of epidemiological research (Sartwell 1955).

But obviously, although certain preventive measures might be studied experimentally, in general, as noted by several of the participants in the discussion, "opportunities for carrying out such studies are rare" (Lilienfeld 1957, p. 52). During the late 1950s and early 1960s, some of the major figures in the emerging field of chronic disease epidemiology therefore discussed in a number of papers how to investigate etiologic factors and especially when to make inferences about causality (Lilienfeld 1957; Yerushalmy and Palmer 1959; Sartwell 1960; Lilienfeld 1959; Wynder and Day 1961; Hill 1965). As argued by Parascandola (Parascandola 2004a; Parascandola 2004b), two different approaches can be recognized in this discussion: a rigorous/formalistic and a pragmatic one. Thus, a number of senior biostatisticians who were eager to protect the status of their methods just as they had become accepted through the RCTs emphasized the rigorous pursuit of basic principles of experimental design and focussed on guidelines intended to minimize the risk of spurious correlations, such as, for example consistency and specificity. On the other side, several of the public health officials were driven by a pragmatic need to implement preventive measures from the data available and stressed such guidelines as, for example, the overall biological plausibility of a causal relation or analogy to similar cases.

By the same token, there were confronting views on the adequacy of a reductionist focus on individual agents. Thus, whereas some emphasized how bacteriology had developed towards "the orderly and systematic identification of causative organisms in many diseases" (Yerushalmy and Palmer 1959, p. 28) and praised the identification of the typhoid bacillus over studies of polluted water and milk (ibid., p. 29), others stressed that even if the knowledge of the causal agent was inconclusive, epidemiological data could still be used for preventive measures and referred to such classic heroes of epidemiology as Snow, Jenner and Farr to vindicate this view (see e.g.(Lilienfeld 1957; Sartwell 1960; Wynder and Day 1961; Wynder 1961).

Since the suggested guidelines expressed rules of thumb rather than logically valid criteria, they all had both their defenders and opponents, and the discussions never reached any clear consensus. For example, with respect to specificity, the formalistically inclined in the discussion saw it as an important guideline to ascertain that "the difference in relative frequencies reflects a specific and meaningful relationship between the characteristic under suspicion and the disease under consideration" (Yerushalmy and Palmer 1959, p. 36). Thus, on their view, if a characteristic was related to several diseases that were not obviously physiologically or pathologically related, "the relationship must be assumed – until further proof – to be nonspecific" (ibid., p. 37) and further studies were needed to judge whether the association carried any meaning or not. On the other hand, the pragmatically inclined in the discussion dismissed this argument that seemed "to rest on the assumption that a particular characteristic can induce only a single disease" (Sartwell 1960, p. 62); an assumption that was easily rejected with reference to such examples as excessive alcohol consumption causing both acute gastritis, cirrhosis of the liver and psychosis (ibid.).

Likewise, while the pragmatically inclined would focus on biological plausibility, claiming that "if a statistical association makes biological sense, it is more readily accepted than one that is at the moment not capable of biological explanation" (Lilienfeld 1957, p. 57), the formalistically inclined would maintain that "the most important consideration with respect to a theory is not whether it appears plausible, but whether it suggests experiments" (Berkson 1958, p. 35f).

Initially, participants in the debate advanced their own favourite lists of criteria or attacked some criteria advanced previously (see Evans 1976; Evans 1978) for an overview). But gradually, lists of criteria were compiled, both in reports from major epidemiological studies to support specific conclusions, and in papers dealing more abstractly with the methodology of chronic disease epidemiology. Thus, the Surgeon General's 1964 report on smoking, *Smoking and Health,* included an introductory chapter on causality in which it was stated that to judge the causal significance of an association "a number of criteria must be utilized, no one of which is an all-sufficient

basis for judgement" (Surgeon General 1964, p. 20). These criteria were listed as consistency, strength, specificity, temporal relationship, and coherence. By the same token, in 1965 an even longer list was compiled by the medical statistician Austin Bradford Hill in his presidential address to the newly-founded Section of Occupational Medicine in the Royal Society of Medicine. In this address he presented a list of nine aspects that should be considered before deciding if a given association was likely to express a causal relation: 1) the strength of the association, 2) the consistency of the observed association, 3) the specificity of the association, 4) temporality, 5) biological gradient, or dose-response curve, 6) biological plausibility. 7) coherence, 8) experiment, and 9) analogy. During the 1960es and 1970es several textbooks would adopt versions of some of these lists, such as (MacMahon and Pugh 1970; Lilienfeld and Lilienfeld 1980).[8] Hill's list came to be by far the most influential, and the nine guidelines would often be presented as 'criteria', parallel to the Henle-Koch criteria. However, Hill himself had been careful to emphasize that "[w]hat I do not believe … is that we can usefully lay down some hard-and-fast rules of evidence that must be obeyed before we accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non" (Hill 1965, p. 299); a warning that was often lost in later presentations of the list. The Hill guidelines thus became a permanent display of the 'split' between the rigorous/formal and the pragmatic approach, with pragmatics on the one hand underscoring their usefulness in structuring the valuation of causation (e.g. Doll 2002), and formalists on the other hand emphasizing how all the nine guidelines, save temporality, "are saddled with reservations and exceptions" (Rothman 1998, p. 27).

[8] See (Zhang 2004) for an overview of the development of epidemiology textbooks, including their treatment of causality.

5. Implicit normativity in the debate

But the longevity and the occasionally implacable tone of the debate was due to more than a split between formal and pragmatic inclinations. Another theme was also implicitly present in the discussions, namely the *consequences* of drawing or not drawing such inferences.

Thus, in the smoking and lung cancer debate, several of the epidemiologists involved in the various studies on smoking argued that hesitance against preventive actions was due to cultural entrenchment and financial interest. Some suggested that "if the findings had been made on a new agent, to which hundreds of millions of adults were not already addicted, and on one which did not support a large industry, skilled in the arts of mass persuasion, the evidence for the hazardous nature of the agent would be generally regarded as beyond dispute" (Cornfield et al. 1959, p. 198). Others emphasized that if even a single case of cancer could be prevented, it was the duty of medical doctors to do so, or that the difficulties curing or just alleviating the chronic diseases caused by tobacco clearly justified the application of preventive measures (Wynder and Day 1961).

Such reflections of the normative issues involved in the acceptance or rejection of causal hypotheses were also included by Hill in his seminal Presidential address. After the presentation of his nine guidelines, Hill added that in areas such as occupational medicine where the ultimate goal was to take action, it was inevitable to introduce what he called "differential standards":

> on relatively slight evidence we might decide to restrict the use of a drug for early-morning sickness in pregnant women. If we are wrong in deducing causation from association no great harm will be done. The good lady and the pharmaceutical industry will doubtless survive. On fair evidence we might take action on what appears to be an occupational hazard, e.g. we might change from a probably carcinogenic oil to a non-carcinogenic oil in a limited environment and without too much injustice if we are wrong. But we should need very strong evidence before we make people burn a fuel in their homes that they do not like or stop smoking the cigarettes and eating the fats and sugar that they do like. (Hill 1965, p. 300).

Thus, on Hill's view, a possible serious effect of a product that could be removed without any negative effects – for example products similar to Thalidomide which in the late 1950s had become a popular treatment of morning sickness, and whose potential for causing congenital malformation of babies born by the women who took the drug had become clear just a few years before Hill's talk[9] - only slight evidence would be needed to implement preventive actions. On the other hand, if it had negative effects to remove a product, for example because the consumers had a strong preference for it, much stronger evidence was needed.

But to others, such as Fisher, the idea of including considerations of the consequences when deciding between hypotheses was anathema. In the statistical literature Neyman and Pearson had also suggested including the consequence of error when deciding between hypotheses (Neyman and Pearson 1933). Introducing the distinction between what has later become known as type I and type II errors, they argued that whether it is more important to avoid rejecting a true hypothesis or accepting a false one depends on the consequences of the error:

> If we reject H_0 we may reject it when it is true; if we accept H_0 we may be accepting it when it is false, that is to say, when really some alternative H_t is true. These two sources of error can rarely be eliminated completely; in some cases it will be more important to avoid the first, in others the second. We are reminded of the old problem considered by Laplace of the number of votes in a court of judges that should be needed to convict a prisoner. Is it more serious to convict an innocent man or to acquit a guilty? That will depend upon the consequences of the error; is the punishment death or fine; what is the danger to the community of released criminals; what are the current ethical views on punishment? (Neyman and Pearson 1933, p. 296)

[9] Ironically, Hill's example should later prove problematic. Thus, the medical doctor McBride who had been among the first to warn against Thalidomide, later warned against another morning sickness drug that he suspected could cause similar birth deformities. The drug was later found safe (Smith 1981; Kolata 1985; Barinaga 1995), but the drug company still felt forced to stop marketing the product (Culliton 1983). It turned out that McBride has only sparse evidence for this claim and that he had falsified data in one of his animal studies of the drug (Lawson 1993).

As described by Howie (2002) and Marks (2003), Fisher was strongly opposed to the Neyman-Pearson approach. He had advanced a theory of statistical inference that enabled a numerical expression of confidence in a hypothesis. Thus, for Fisher, the important point was the search for truth, whereas the utility of a hypothesis should not enter the investigation of its truth. The Neyman-Pearson approach, on the contrary, was a strategy for deciding between different courses of action, that is, it was not as such concerned with whether a hypothesis was true, but whether one should act as if it was (Howie 2002, p. 178). Thus, in explicit opposition to Neyman and Pearson, Fisher emphasized the important difference between significance tests as a means to the improvement of knowledge and their interpretation into "that technological and commercial apparatus which is known as an acceptance procedure" (Fisher 1955, p. 69), adding that

> I am casting no contempt on acceptance procedures, and I am thankful, whenever I travel by air, that the high level of precision and reliability required can really be achieved by such means. But the logical differences between such an operation and the work of scientific discovery by physical or biological experimentation seem to me so wide that the analogy between them is not helpful, and the identification of the two sorts of operation is decidedly misleading (Fisher 1955), p. 69f).

As argued by Marks (2003), although Fisher's dissatisfaction was originally based on statistical methodology, he later developed a political critique of the idea that the realm of science and the realm of decisions were the same (cf. Marks 2003, p. 934); a confusion that he thought Neyman had imported from Eastern Europe where Fisher felt that decisions about science were a matter of state policy.[10] In contrast, Fisher wanted to defend 'the right of other free minds' to make 'their own decisions' (cf. Marks 2003, p. 934).

[10] Letter from Fisher to Keyfitz, 21. November 1955, cf Marks 2003, p. 934.

This insistence on a distinction between science and application also seemed to be shared by the epidemiologists who saw it as a distinction between the pure scientist occupied only with truth and the health officer occupied with actions to protect the population. Thus, also Hill emphasized that

> in passing from association to causation I believe in 'real life' we shall have to consider what flows from that decision. On scientific grounds we should do no such thing. The evidence is there to be judged on its merits and the judgment (in that sense) should be utterly independent of what hangs upon it – or who hangs because of it. But in another and more practical sense we may surely ask what is involved in our decision (Hill 1965, p. 300).

However, at a closer look this distinction was not that clear. Lilienfeld, for example, in his seminal paper on epidemiological methods and inferences also assumed a difference between the research scientist and the health officer, but it was a difference in the degree of required evidence to accept a hypothesis rather than a fundamental difference between truth and action:

> Certain nonbiological considerations may influence an individual's attitude toward acceptance of a causal inference. These concern the decisions that are made relative to the course of action to be taken when an inference is accepted. They reflect the outlook, background, and administrative responsibilities of the individual. For example, a research scientist, without any direct responsibility for the health of a population, might require a very high degree of plausibility before accepting a causal inference and recommending definite administrative action. On the other hand, a health officer, directly responsible for the health of a population, may accept a lower degree of plausibility as sufficient to warrant preventive action. (Lilienfeld 1957, p. 58)

What was implicitly revealed here was that the idea of the ideal scientist concerned solely with truth and isolated from any impure considerations of actions was a phantom. Even the research scientist might eventually become involved in 'recommending definite administrative action'. Looking closely at the debate on tobacco, Fisher's stand was not solely to refrain from recommendations. In some of his publications he

clearly recommended *not* to take preventive measures, arguing that such measures would interfere "with the peace of minds and habits" of the population and "plant fear in the minds of perhaps a hundred million smokers" (Fisher 1958b, p. 12), maybe even create "states of frantic alarm" (Fisher 1957b, p. 43). Thus, when he argued that "statistics has gained a place of modest usefulness in medical research. It can deserve and retain this only by complete impartiality, which is not unattainable by rational minds" (Fisher 1957a, p. 298) he overlooked that a decision to await further research before implementing preventive action was also a normative decision, just driven by other norms than the contrary decision to implement preventive action on the available evidence.

Whereas Hill's examples had pointed to the seriousness of the putative effect, the (in)dispensability of the product, personal preferences and the proximity to the private sphere of the consumers as some of the key values to enter into the decision, Fisher was primarily concerned with the protection of the private sphere and the individual autonomy; a view that was also put forward not only by the tobacco industry, but also in some editorials where, for example, the New England Journal of Medicine in 1961 emphasized that "Each individual must choose his own course, whether to woo the lady nicotine or abjure the filthy weed, while the search for truth continues" (anon. 1961).[11]

---

[11] Only rarely was it described how the individual was supposed to choose his or her own decision. One of the only such accounts, provided by the physicist Brownlee in response to the Surgeon General's report, seems to assume a fully rational agent capable of embracing all available evidence and having a clear weighing of life expectancy, pleasures and nuisances: "The individual smoker would than have to weigh the above changes in expectation by his assessment of the question of whether the cigarette smoking was a cause. He would have to relate this expected change in expectation of life to the analogous probable expected changes in expectation of life caused by various other of his activities which probably cause decreases in expectation of life, such as over-eating, undersleeping, underexercising, travelling by automobile, etc. He would also have to weigh the possibility of such adverse consequences of giving up smoking as in due cause becoming appreciably overweight. He would then have to assign utilities to the expected change in expectation of life and to the pleasure and satisfaction he gets from cigarette smoking" (Brownlee 1965), p. 738). Brownlee had worked primarily on statistical quality control and may therefore have been acquainted with the decision theories used in economics at the time. However, his imagined calculus is far from the standards in the medical community at the time where emphasis until the 1970s was more on paternalism than on autonomous patients embracing all available evidence in forming their own decisions. Further, when an explicit pub-

Although these normative issues were behind much of the strong rhetoric, the normative character of the debate was not recognized.[12] Instead, the debate was perceived as primarily concerned with methodology. However, the methodological question – to provide criteria for how to draw inferences about causality in multicausal settings – was in principle unsolvable and the debate therefore never reached an actual closure but continued, for example through the recurrent praise and scorn, respectively, of Hill's guidelines. Further, focusing on this unsolvable methodological issue, it was never explicitly discussed which values were at stake and it appeared as a matter of taste rather than the result of normative analysis whether or not to make public recommendations from epidemiological evidence.

References

anon. Experimental Links Between Tobacco and Lung Cancer. BMJ , 1050-1051. 1958.
anon. The Great Debate (editorial). New England Journal of Medicine 254, 1266. 1961.
anon. The Ciagarette as Co-carcinogen. Lancet 279, 85-86. 1962.
Arkin, H. Relationship between human smoking habits and death rates. Current Med.Digest 22, 37-44. 1955.
Barinaga, M. Bendectin Case Dismissed. Science 267, 167. 1995.
Bayer, R. and Fairchild, A. L. The genesis of public health ethics. Bioethics 18[6], 473-492. 2004.
Berkson, J. Limitations of the Application of Fourfold Table Analysis to Hospital Data. Biometrics Bulletin 2[3], 47-53. 1946.
Berkson, J. The statistical study of association between smoking and lung cancer. Proc Staff Meet Mayo Clin 30[15], 319-348. 1955.
Berkson, J. Smoking and Lung Cancer: Some Observations on Two Recent Reports. Journal of the American Statistical Association 53, 28-38. 1958.
Berlivet, L. 2005. Association or Causation? The Debate on the Scientific Status of Risk Factor Epidemiology, 1947- c.1965. In *Making Health Policy. Networks in Research and Policy after 1945.* Edited by V. Berridge. Amsterdam: Rodopi.
Brandt, A. M. The Cigarette, Risk, and American Culture. Daedalus 119[4], 155-176. 1990.
Carter, K. Codell. Koch's Postulates in Relation to the Work of Jacob Henle and Edwin Klebs. Medical History 29, 353-374. 1985.

---

lic health ethics finally began to be developed, it was explicitly questioned whether a priority of individual autonomy could be assumed to be appropriate (Bayer and Fairchild 2004).

[12] As Berridge has argued, there has been a tendency in the historical literature to base interpretations of the history of the smoking debate on the assumption of value-free scientific truth and the notion of progress towards a correct cause of action (Berridge 2006), p. 1188). This corresponds with the distinction between science and application assumed by the historical actors and contributes to hiding the inherently value-based character of the debate.

Carter, K. Codell. The Development of Pasteur's Concept of Disease Causation and the Emergence of Specific Causes in Nineteenth-Century Medicine. Bulletin of the History of Medicine 65, 528-548. 1991.

Cornfield, H., Haenzel, W., Hammond, E. C., Lilienfeld, AM., Shimkin, M. B., and Wynder, E. L. Smoking and lung cancer. recent evidence and a discussion of some questions. Journal of the National Cancer Institute 22, 173-203. 1959.

Culliton, B. J. Merell Dow Stops Marketing Bendectin. Science 221, 37. 1983.

Dawber, T. R. 1980. *The Framingham study: the epidemiology of coronary heart disease.* Cambridge MA: Harvard University Press.

Dean, H. T. 1992. Fluoridation: Success and Controversy. In *Dental history in a new age: A history of the National Institute of Dental Research.* Edited by R. R. Harris. Ames: Iowa State University Press.

Dean, H. T., Arnold, F. A., Jay, P., and Knutson, J. W. Studies on Mass Control of Dental Caries Through Flouridation of the Public Water Supply. Public Health Reports 65, 1403-1408. 1950.

Doll, R. Uncovering the effects of smoking: historical perspective. Statistical Methods in Medical Research 7[2], 87-117. 1998.

Doll, R. Proof of causality. Deduction from epidemiological observation. Perspectives in Biology and Medicine 45[4], 499-515. 2002.

Doll, R. Science wil always win in the end. Interview by Anna Wagstaff. Cancer World December 2004, 28-34. 2004.

Doll, R. and Hill, A. B. Smoking and carcinoma of the lung. BMJ 2(4682), 739-748. 1950a.

Doll, R. and Hill, A. B. Lung Cancer and Other Causes of Death in Relation to Smoking. BMJ 2(5001), 1071-1081. 1956.

Doll, R. and Hill, A. B. Smoking and carcinoma of the lung. BMJ , 739-748. 1950b.

Evans, A. S. Causation and disease: A chronological journey. American Journal of Epidemiology 108, 249-258. 1978.

Evans, A. S. Causation and Disease: The Henle-Koch Postulates Revisited. Yale Journal of Biology and Medicine 49, 175-195. 1976.

Fisher, R. A. Statistical Methods and Scientific Induction. Journal of the Royal Statistical Society 17[1], 69-78. 1955.

Fisher, R. A. Dangers of cigarette smoking. BMJ 2, 297-298. 1957a.

Fisher, R. A. Cancer and smoking? (letter). Nature 182, 596. 1958a.

Fisher, R. A. Cigarettes, Cancer, and Statistics. Centennial Review 11[2], 151-166. 1958b.

Fisher, R. A. Lung cancer and cigarettes? (letter). Nature 182, 108. 1958c.

Fisher, R. A. 1959. *Smoking. The Cancer Controversy.* Edinburgh: Oliver and Boyd.

Fisher, R. A. 1935. *Design of Experiment.* Edinburgh: Oliver and Boyd

Fisher, R. A. Dangers of Cigarette-smoking. BMJ 2[43], 43. 1957b.

Gordon, T. and W. Kannel. 1970. The Famingham,Massachusetts study - 20 years later. In *The community as an epidemiologic laboratory: a casebook of community studies.* Edited by Kessler and M. L. Levin. Baltimore: Johns Hopkins University Press.

Hammond, E. C. and Horn, D. Smoking and death rates; report on forty-four months of follow-up of 187,783 men. I. Total mortality. JAMA 166[10], 1159-1172. 1958a.

Hammond, E. C. and Horn, D. Smoking and death rates: report on forty-four months of follow-up of 187,783 men. 2. Death rates by cause. JAMA 166[11], 1294-308. 1958b.

Hill, A. B. The Clinical Trial. BMJ , 278-282. 1951.

Hill, A. B. The Clinical Trial. New England Journal of Medicine 247, 113-119. 1952.

Hill, A. B. Observations and Experiment. New England Journal of Medicine 248, 995-1001. 1953.

Hill, A. B. The Environment and Disease: Association or Causation? Proceedings of the Royal Society of Medicine 58, 295-300. 1965.

Hill, A. B. and Doll, R. Lung Cancer and Tobacco. The BMJ's Questions Answered by Professor A. Bradford Hill and R. Doll. BMJ 1(4976), 1160-1163. 1956.

Hill, G., Millar, W., and Connelly, J. "The Great Debate": Smoking, Lung Cancer, and Cancer Epidemiology. Canadian Bulletin of Medical History 20[2], 367-386. 2003.

Howie, D. 2002. *Interpreting Probability. Controversies and Developments in the Early Twentieth Century.* Cambridge: Cambridge University Press.

Kluger, R. 1996. *Ashes to ashes : America's hundred-year cigarette war, the public health, and the unabashed triumph of Philip Morris.* New York: Alfred A. Knopf.

Kolata, G. B. Jury Clears Bendectin. Science 227, 1559. 1985.

Lawson, M. McBride found guilty of fraud. Nature 361, 673. 1993.

Levin, M. L., Goldstein, H., and Gerhardt, P. R. Cancer and tobacco smoking; a preliminary report. JAMA 143[4], 336-338. 1950.

Lilienfeld, AM. "On the methodology of investigations of etiologic factors in chronic dieseases" - some comments. Journal of Chronic Diseases 10[1], 41-46. 1959.

Lilienfeld, AM. Epidemiological Methods and Inferences in Studies of Noninvectious Diseases. PHR Review 72[1], 51-60. 1957.

Lilienfeld, AM. and D. E. Lilienfeld. 1980. *Foundations of epidemiology.* New York: Oxford University Press.

Little, C. C. Some phases of the problem of smoking and lung cancer. New England Journal of Medicine 264, 1241-1245. 1961.

Lock, S., L. A. Reynolds, and E. M. Tansey. 1998. *Ashes to ashes: The history of smoking and health.* Amsterdam: Rodopi.

MacMahon, B. and T. F. Pugh. 1970. *Epidemiology; principles and methods.* Boston: Little & Brown.

MacMahon, B., T. F. Pugh, and J. Ipsen. 1960. *Epidemiologic Methods.* Boston: Little, Brown & Co.

Marks, H. M. Rigorous uncertainty: why RA Fisher is important. International Journal of Epidemiology 32, 932-937. 2003.

Morris, J. N. 1964. *Uses of Epidemiology.* Edinburgh: E. & S. Livingstone.

Neyman, J. and Pearson, E. S. On the Problem of the Most Efficient Tests of Statistical Hypotheses. Philosophical Transactions of the Royal Society of London.Series A, Containing Papers of a Mathematical or Physical Character 231, 289-337. 1933.

Paneth, N., Susser, E., and Susser, M. Origins and early development of the case-control study: part 1, Early evolution. Soz.-Präventivmed. 47, 282-288. 2002a.

Paneth, N., Susser, E., and Susser, M. Origins and early development of the case-control study: part 2, The case-control study from Lane-Claypon to 1950. Soz.-Präventivmed. 47, 359-365. 2002b.

Parascandola, M. Epidemiology: Second-rate science? Publich Health Reports 113, 312-320. 1998.

Parascandola, M. Skepticism, Statistical Methods, and the Cigarette. Perspectives in Biology and Medicine 47, 244-261. 2004a.

Parascandola, M. Two approaches to etiology: the debate over smoking and lung cancer in the 1950s. Endeavour 28[2], 81-86. 2004b.

Rothman, D. J. 1991. *Strangers at the Bedside: A History of How Law and Bioethics Transformed Medical Decision Making.* New York: Basic Books.

Rothman, K. J. 1998. *Modern Epidemiology.*Lippincott Williams & Wilkins.

Sartwell, P. E. "On the methodology of investigations of etiologic factors in chronic diseases" - Further comments. Journal of Chronic Diseases , 61-63. 1960.

Sartwell, P. E. Some Approaches to the Epidemiologic Study of Chronic Disease. Am.J.Public Health 45, 609-614. 1955.

Smith, R. J. Studies Support Bendectin Safety Claim. Science 212, 1485. 1981.

Stolley, P. D. When Genius Errs: R.A. Fisher and the Lung Cancer Controversy. American Journal of Epidemiology , 416-425. 1991.

Summers, W. C. Causal theories, justification and experimental science: Daniel E. Salmon and hog cholera. 2007.

Surgeon General. 1964. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service.* Washington: US Department of Health, Education, and Wealfare.

Susser, M. Epidemiology in the United States after World War II: The Evolution of Technique. Epidemiologic Reviews 7, 147-177. 1985.

Susser, M. Choosing a Future for Epidemiology: I. Eras and Paradigms. Am.J.Public Health 86, 668-677. 1996a.

Susser, M. Choosing a Future for Epidemiology: II. From Black Box to Chinese Boxes and Eco-Epidemiology. Am.J.Public Health 86, 674-677. 1996b.

Talley, C., Kuschner, H. I., and Sterk, C. E. Lung Cancer, Chronic Disease Epidemiology, and Medicine, 1948-1964. Journal of the History of Medicine and Allied Sciences 59[3], 329-374. 2004.

Vandenbroucke, J. P. The history of confounding. Soz.-Präventivmed. 47, 216-224. 2002.

Vineis, P. History of bias. Soz.-Präventivmed. 47, 156-161. 2002.

Wynder, E. L. An appraisal of the smoking-lung-cancer issue. New England Journal of Medicine 264, 1235-1240. 1961.

Wynder, E. L. Tobacco as a Cause of Lung Cancer: Some Reflections. American Journal of Epidemiology 146[9], 687-694. 1997.

Wynder, E. L. and Day, E. Some Thoughts on the Causation of Chronic Disease. JAMA 175[11], 997-999. 1961.

Wynder, E. L. and Graham, E. A. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases. JAMA 143[4], 329-336. 1950.

Wynder, E. L., Graham, E. A., and Croninger, A. B. Experimental Production of Carcinoma with Cigarette Tar. Cancer Research . 1953.

Yerushalmy, J. and Palmer, C. E. On the methodology of investigations of etiologic factors in chronic diseases. Journal of Chronic Diseases 10, 27-40. 1959.

Zhang, F. F. et al. Evolution of epidemiologic methods and concepts in selected textbooks in the 20th century. Soz.-Präventivmed. 49, 97-104. 2004.