9. Causal inference[*]

The desire to act on the results of epidemiologic studies frequently encounters vexing difficulties in obtaining definitive guides for action. Weighing epidemiologic evidence in forming judgments about causation.

"The world is richer in associations than meanings, and it is the part of wisdom to differentiate the two." — John Barth, novelist.

"Who knows, asked Robert Browning, but the world may end tonight? True, but on available evidence most of us make ready to commute on the 8.30 next day." — A. B. Hill

Historical perspective

Our understanding of causation is so much a part of our daily lives that it is easy to forget that the nature of causation is a central topic in the philosophy of science and that in particular, concepts of disease causation have changed dramatically over time. In our research and clinical practice, we largely act with confidence that 21st century science has liberated us from the misconceptions of the past, and that the truths of today will lead us surely to the truths of tomorrow. However, it is a useful corrective to observe that we are not the first generation to have thought this way.

In the 1950s, medical and other scientists had achieved such progress that, according to Dubos (1965:163), most clinicians, public health officers, epidemiologists, and microbiologists could proclaim that the conquest of infectious diseases had been achieved. Deans and faculties began the practice of appointing, as chairs of medical microbiology, biochemists and geneticists who were not interested in the mechanisms of infectious processes. As infectious disease epidemiology continues to surge in popularity, we can only shake our heads in disbelief at the shortsightedness of medical and public health institutions in dismantling their capabilities to study and control infectious diseases, whose epidemics have repeatedly decimated populations and even changed the course of history.

On the other hand, perhaps the connection is not so direct. According to Dubos (1965), the 19th-century fall in death rates from infectious disease and malnutrition actually began in mid-century, several decades before the medical discoveries of the scientific era could be turned into actual policies. Medical science and the germ theory received an inordinate share of credit because the decline was not widely recognized until the end of the century. Moreover, he charges

> The present generation [presumably the pre-World War II birth cohorts going back to 1910] goes still further and now believes that the control of infectious and nutritional disease dates from the widespread use of antibacterial drugs and from the availability of vitamins and processed foods. So short and parochial are our memories!" (page 365)

* (An earlier version of these notes was prepared by Sandra Martin, Ph.D.)

While acknowledging the very important roles played by local boards of health and other municipal bodies, Dubos (1965) attributes most of the improvement in health to improvements in prosperity and transportation that enabled many people to affort "at least one square meal a day":

> No medical discovery made during recent decades can compare in pratcial importance with the introduction of social and economic decency in the life of the average man. The greatest advances in the health of the people were probably the indirect results of better housing and working conditions, the general availability of soap, of linen for underclothing, of glass for windows, and the humanitarian concerns for higher living standards. (page 365)

Before proceeding to our investigation of causal inference, it will be helpful to take a brief look at the history of public health and disease in the 17th-19th centuries. [The following account comes primarily from Wilson Smillie (*Public health: its promise for the future*), Mervyn Susser (*Causal thinking in the health sciences*), and Lisa Berkman and Lester Breslow (*Health and ways of living*).]

In the early seventeenth century, medical science was "just emerging from the morass of the Middle Ages" (Smillie, 1955:18). The most deadly disease of the American colonies in that century was smallpox. The disease was even more devastating to the indigenous populations of the New World and is believed to have killed over half the Indian population in Mexico following the Spanish conquest (Smillie, 1955:21). In Europe, smallpox was an endemic disease of childhood; but in the more isolated situation in the colonies, recurrent epidemics devastated settlements. According to Smillie (p22), a 1633 smallpox epidemic in the Massachusetts Bay colony spread to the Indians and killed "whole plantations" of them. A 1689-1690 epidemic in New England killed 1,000 people in one year (by comparison, Boston had a total population of about 7,000) at that time.

During the eighteenth century, the practice of smallpox inoculation (published by the Greek Timonius in 1714) successfully aborted epidemics in the American colonies, although the practice was at first resisted. Smallpox inoculation was banned outright in New York City in 1747, required the governor's permission in Carolina in 1764, and required the consent of the selectmen in towns in Massachusetts (Smillie, 1955, p28).

Nevertheless, smallpox inoculation soon proved its worth. At the beginning of the Revolutionary War, in 1776, smallpox arrived in Boston. A heroic campaign inoculated 9,152 nonimmune people in three days. Although the inoculations produced 8,114 smallpox cases resulting in 165 deaths (1.8%), the 232 natural infections in susceptible persons who had not been inoculated resulted in 33 deaths (14.2%) (data from Shattuck, Lemuel, reported in Smillie, 1955:29). Two decades later, Edward Jenner, an obscure country practitioner in England, demonstrated immunity to smallpox in ten persons who had previously developed cowpox. Although his paper to the Royal Society was refused, he published his classic monograph in 1798 to become known as the father of vaccination. (Take-home message: don't let a manuscript rejection discourage you!)

The second great communicable disease in eighteenth century North America, yellow fever, also took a fearsome toll on communities in the New World. The first American article on yellow fever (by Dr. John Lining of Charleston) described the disease as both contagious and imported (Smillie,

1955:35). Quarantine of sick persons and of ships suspected of having yellow fever on board was sometimes instituted to prevent or abort epidemics.

Towards the end of the eighteenth century, though, the miasma theory of disease arose – the theory that all disease was due to bad air – contaminations (miasma) emanating from a great variety of sources (Smillie, 1955:3). So strong was the power of this new theory, that Dr. Benjamin Rush, the greatest American physician of that time, was sure that the great 1793 yellow fever epidemic in Philadelphia was neither contagious nor had come from the West Indies, but was rather due to a pile of spoiled coffee thrown on a wharf (Smillie, 1955:9).

By the early nineteenth century, medicine and the public health movement were dominated by the miasma theory (Susser, *Causal Thinking in the Health Sciences*). The line of investigation was to prove the ill-effects of miasma; the line of prevention was to eliminate the sources of miasma in slums and poor sanitation. Although the concept of miasma, overthrown later in the century, is ridiculed today, the sanitation measures that the miasma theory called for were often dramatically effective in reducing death rates. During the nineteenth century, as Susser writes, Jacob Henle formulated the conditions that needed to be met to prove the germ theory, and some 20 years later, Louis Pasteur demonstrated the existence of microorganisms. Now, the causes of disease could actually be seen – microbiology had progressed from a science of inference to a science of direct observation. Microorganisms then became the object of the search for causes. The containment of the spread of microbes became the object of prevention. Asepsis, antisepsis, and disinfection - measures taken on the basis of germ theory - also proved effective. Moreover, the new paradigm proved superior to the miasma theory through its greater specificity and in its ability to explain and predict certain phenomena outside the miasma theory, such as immunization and chemotherapy.

The discovery of microorganisms and the ascendence of the germ theory of disease reinforced the view that illness consisted of many discrete clinical entities, each caused by a different agent, and each with certain morbid manifestations yielding distinct syndromes (Berkman and Breslow, 1983:5; see also Dubos, 1965). This concept prevails even today, as illustrated in the dictionary definitions shown in the chapter on the Phenomenon of Disease. The search for specific agents has led to great breakthroughs in medicine and public health, such as the effective control of many infectious diseases in the developed world and the worldwide eradication of smallpox. Even where the germ theory did not apply, as in the case of vitamin deficiency diseases, the concept of specificity of cause has also proved effective for etiology and control.

There were those who resisted the one-disease-one-cause model. But the tide was against them. As Dubos (quoted in Berkman and Breslow, p. 6) observed:

> These vague arguments were no match for the precise experimentation by which Pasteur, Koch, and their followers defended the doctrine of specific causation of disease. Experimental science triumphed over the clinical art, and within a decade the theory of specific etiology of disease was all but universally accepted, soon becoming, as we have seen, the dominant force in medicine.

At the same time, this great spurt in medical research diminished awareness of the rarity of one-to-one relationships and of the complex relationships between causes and effect that exist in the real

world. Even as late as the 1950's, for example, it was very difficult to conceptualize that smoking can cause so many diseases (noted by the late Ernst Wynder in a 1997 seminar at UNC at Chapel Hill), and the fact that so many diseases were associated with cigarette smoking was put forward as an argument against interpeting the associations as causal. (According to Sir Richard Doll, in 1992 the Sloan-Kettering Institute's new director "told Wynder that his conclusiion that a causal relationship existed between smoking and lung cancer was irresponsible and that all future publications by his goup would have to be cleared through the director's office", Ernst Wynder, 1923-1999, *AJPH* 1999;89:1798-9: 1799).

Lord Bertrand Russell has written, "every advance in a science takes us further away from the crude uniformities which are first observed into a greater differentiation of antecedent and consequent and into a continually wider circle of antecedents recognized as relevant" (*Mysticism and Logic*, London: Longmans, Green, 1918, p. 188, quoted in E.H. Carr, *What is history*, NY: Knopf, 1963, p. 118). A number of developments undermined the supremacy of the one-cause-one-disease model.

One was the growing predominance of microbial diseases of endogenous origin, diseases caused by organisms that are carried by many people in the population (Dubos, 1965:176). Contempory examples are bacterial infections secondary to acute viral illnesses, opportunistic infections in persons with AIDS, and urinary tract infections with *E. coli*. A second was the recognition that many pathogens, including the tubercle bacillus, can be carried for long periods of time, only to cause disease when the host's immunity becomes weakened. A third was the shift of attention from infectious diseases to heart disease and cancer, where various factors are related to risk but none absolutely necessary; thus the term "multifactorial" disease. (Although CHD is the classic multifactorial disease, there have been recent suggestions that infectious processes may be an important dimension.) Finally, as epidemiology has expanded to study behavioral and environmental maladies (e.g., automobile injuries, alcoholism, homicide, and unprotected intercourse), a unicausal model does not even have meaning.

Nevertheless, in practice much of epidemiology focuses on single risk factors. Ideally we could make use of an overall model combining multiple etiologic agents into a comprehensive system. But often epidemiologic research has its greatest role in stages of investigation before a comprehensive causal picture is possible. Indeed, epidemiologic studies are one of the primary avenues towards beginning to define the factors that might make up such a picture. So the usual approach is to take one or two suspected factors at a time and then see if, taking into account what has already been discovered about the disease, the suspected factors increase the explanatory or predictive power of the investigation. This one factor-at-a-time approach is the essence of risk factor epidemiology, of the concepts confounding and effect modification to be presented later, and of epidemiologic approaches to causal inference.

The concept of causality

In *Modern Epidemiology*, Rothman and Greenland illustrate the process of understanding a cause with a description of a toddler learning that moving a light switch causes the light to turn on. But what we take as a cause depends upon the level at which we seek understanding or the constituency we represent. Thus:

The **mother** who replaced the burned-out light bulb may see her action as the cause for the light's turning on, not that she denies the effect of the light switch but has her focus elsewhere.

The **electrician** who has just replaced a defective circuit breaker may cite that as the cause of the light's turning on, not that he denies the importance of the switch and the bulb, but his focus is elsewhere still.

The **lineman** who repaired the transformer that was disabled by lightning may regard his repair as the cause of the light's turning on.

The **social service agency** that arranged to pay the electricity bill may regard that payment as the cause of the light's turning on, since with the electricity cut off, neither the switch nor the circuit breaker matters.

The **power company**, the **political authority** awarding the franchise, the investment bankers who raised the financing, the **Federal Reserve** that eased interest rates, the **politician** who cut taxes, and the **health care providers** who contributed to the toddler's safe birth and healthy development might all cite their actions as the real cause of the light's turning on.

The National Rifle Association's slogan "Guns don't kill people; people kill people" is not a public health stance, but it does illustrate the complexities of apportioning causation.

Mervyn Susser proposes that for epidemiologists a causal relation has the following attributes: association, time order, and direction. A cause is something that is associated with its effect, is present before or at least at the same time as its effect, and acts on its effect. In principle, a cause can be **_necessary_** – without it the effect will not occur – and/or **_sufficient_** – with it the effect will result regardless of the presence or absence of other factors. In practice, however, it is nearly always possible to conceive of other factors whose absence or presence could avert an effect since, as with the light switch example above, assumptions are always present. A fall from a five story building would appear to be a sufficient cause of death. But it could be argued that death would not have resulted had there been a safety net below!

Rothman has elaborated a component causes model that attempts to accomodate the multiplicity of factors that contribute to the occurrence of an outcome. In his model, a sufficient cause is represented by a complete circle (a "causal pie"), the segments of which represent component causes. When all of the component causes are present, then the sufficient cause is complete and the outcome occurs. There may be more than one sufficient cause (i.e., circle) of the outcome, so that the outcome can occur through multiple pathways. A component cause that is a part of every sufficient cause is a necessary cause. The induction period for an event is defined in relation to each particular component cause, as the time required for the remaining component causes to come into existence. Thus, the last component cause has an induction period of zero. This model is useful for illustrating a number of epidemiologic concepts, particularly in relation to "synergism" and "effect modification", and we will return to it in a later chapter.

Causal Inference

Direct observation vs. inference:

Much scientific knowledge is gained through direct observation. The introduction of new technology for observation along optical, aural, and chemical dimensions of perception, through such tools as microscopes, x-rays, ultrasound, magnetic resonance scans, and biochemical assays has greatly expanded our opportunities for direct observation and contributed to major advances in scientific knowledge. For example, a recent Nobel Prize was awarded for measuring ion channels in cells, a process that previously had to be inferred. With direct observation, it is possible to "see" causation, especially if one can manipulate the process. Thus, it has been said that the advances in molecular biological techniques have been converting the science of genetics from one of inference to one of direct observation.

In general, however, challenges to understanding transcend that which can be observed directly, so that inference is an essential aspect of scientific activity. It is typically not possible to observe all aspects of a phenomenon of interest, and this situation is very much the case for relationships under epidemiologic investigation. Moreover, even observation involves inference.

Consider the difficulties that arise from latency and induction. The rapidity with which scurvy improved after Lind began his treatments was a great aid in recognizing the effect of lemons. The two-week induction period of measles and its infectiousness before the appearance of symptoms must at one time have been a barrier to understanding its transmission. At the time of Goldberger's investigations, pellagra typically developed about four months after the onset of a niacin-deficient diet. The longer induction period must have made it that much more difficult to associate cause with effect. For example, an interval of four months confounded the seasonality, so that cases were higher in spring and summer (when food was becoming more available) than in winter (when the disease was really developing). At times, opponents of the acceptance of the causal relationship between tobacco and lung cancer have pointed to low rates of lung cancer in populations with high smoking rates (for example, American women in the 1950's) as contradictory evidence, neglecting to take into account the long interval between the onset of cigarette smoking and the development of lung cancer.

Similarly, rare diseases require observation of many subjects, greatly restricting the level of detail that can be visualized or examined. Severe constraints on measurement are also imposed by the need to rely largely on noninvasive measurement methods.

Idealized view of the scientific process

For reasons such as these, a primary recourse in epidemiology is to inference, through:

- positing of conceptual models (conceptual hypotheses);

- deduction of specific, operational hypotheses; and

- testing of operational hypotheses.

As presented in Kleinbaum, Kupper, and Morgenstern, the cycle of scientific progress proceeds as follows:

- Positing of conceptual hypotheses

- Deduction of specific study hypotheses

- Design of study and collection of data

- Analysis of data and conclusions about the study hypotheses

- Modification of the conceptual hypotheses if necessary

This admittedly idealized portrait appropriately emphasizes the importance of conceptual models. As the distinguished historian Edward Hallett Carr has written (*What is history*, NY: Knopf, 1968, p. 136) "The world of the historian, like the world of the scientist, is not a photographic copy of the real world, but rather a working model which enables him more or less effectively to understand it and to master it. The historian distils from the experience of the past, or from so much of the experience of the past as is accessible to him, that part which he recognizes as amenable to rational explanation and interpretation, and from it draws conclusions which may serve as a guide to action. A recent popular writer, speaking of the achievements of science, refers graphically to the processes of the human mind which, 'rummaging in the ragbag of observed "facts," selects, pieces, and patterns the underline relevant observed facts together, rejecting the underline irrelevant, until it has sewn together a logical and rational quilt of "knowledge"' (Leslie Paul: *The annihilation of man*. London: Faber & Faber, 1944, p. 147)."

Carr continues, in a passage that applies much more broadly than to historical reasoning alone, "History therefore is a process of selection in terms of historical significance. To borrow Talcott Parson's phrase once more, history is 'a selective system' not only of cognitive but of causal orientations to reality. Just as from the infinite ocean of facts the historian selects those which are significant for his purpose, so from the multiplicity of sequences of cause and effect he extracts those, and only those, which are historically significant; and the standard of historical significance is his ability to fit them into his pattern of rational explanation and interpretation. Other sequences of cause and effect have to be rejected as accidental, not because the relation between cause and effect is different, but because the sequence itself is irrelevant. The historian can do nothing with it; it is not amenable to rational interpretation, and has no meaning either for the past or the present." (E.H. Carr, *op.cit.*, p. 138). Thus in a hypothetical situation Carr (p. 137) presents in which Jones, driving from a party where he has drunk too much, in a car whose brakes are defective, at an intersection with poor visibility runs down and kills Robinson, who was crossing the road to buy cigarettes, we would entertain alcohol, defective brakes, and poor visibility as causes (and potential targets for preventive action), but not cigarette smoking even though it is true that had Robinson not been a cigarette smoker he would not have been killed that evening.

Conceptual hypotheses arise from underline inductive reasoning, based on available observations and theory, analogies to known processes, and so forth. For example, the effects of passive smoking on lung cancer and of oral contraceptives on breast cancer were first posited based on knowledge of the effects of active smoking on lung cancer and of oral contraceptives on estrogen-sensitive tissues. Existing knowledge may be compatible with more than one alternative model. For example, existing

data on the effects of radiation on cancer risk are compatible with a linear relationship, in which there is no threshold below which risk is absent, or with a curvilinear model in which a threshold of risk exists.

From these conceptual hypotheses, deductive reasoning can generate specific predictions or study hypotheses which ought to be true if the conceptual model is correct. If these predictions or study hypotheses are incompatible with valid data from empirical studies, then the conceptual model that gave rise to the predictions is called into question. This situation forces a re-appraisal or modification of the conceptual hypotheses and lays the basis for advancing understanding.

Karl Popper – the power of falsification:

This aspect of the process of scientific investigation has been emphasized by the philosopher Karl Popper. In Popper's conceptualization, falsification of a hypothesis appears to be more informative than corroboration of a hypothesis. There could be innumerable data sets that are consistent with a false hypothesis. A single counter example, however, forces a modification. Therefore, in Popper's view, studies should _attempt to refute_, rather than to confirm, hypotheses being entertained. A hypothesis that has survived numerous attempts to refute it gains in strength more than one that has merely been corroborated repeatedly.

Although Popper's model is appealing, how well does it describe how science actually proceeds? One problem with this orderly process of induction-deduction-testing requires a large body of knowledge from which to conceptualize and deduce. Particularly in the early stages of research in an area, there is typically a need for descriptive investigations to generate a body of data that can give some direction to thinking about the issues and provides some basis for inductive reasoning. More serious is the fact that in epidemiologic research, a negative result (finding of no association) often cannot refute the original hypothesis because of the many sources of bias that work towards masking underlying associations.

A further point at which the orderly progression outlined above is inadequate is the situation in which the existing conceptual models have been found wanting yet no new ones have been advanced to break through the stalemate. In physics, for example, Einstein's theory of relativity – a revolutionary reconceptualization of physical phenomena – broke through an impasse that had been reached toward the latter part of the 19th century, and opened the way for dramatic advances in knowledge. Goldberger's investigations of pellagra provide a less dramatic but important illustration of the role of a reconceptualization in studying a specific disease. So it is important to bear in mind that advances in knowledge can come from careful observation, precise description, and creative thinking – though in many cases this thinking proceeds through the implicit positing of hypothesis and testing them against available knowledge. Indeed, even the process of direct observation involves paradigms that guide our observation and interpretation.

According to D.C. Stove ("Karl Popper & the Jazz Age"), Popper's philosophy of science can be understood only in reference to the social circumstances of its origins (Vienna in the years after the First World War). In Stove's view, Popper's philosophy is based on reversal of traditional notions of science and philosophy. Traditionally, propositions in science are verifiable. For Popper, they are distinguished by being falsifiable. The method of science has been regarded as essentially inductive.

Popper maintains that it is fundamentally <u>deductive</u>. To many, the essence of science is <u>caution</u>; Popper says that <u>audacity</u> is the essence of science. Science was supposed to be distinguished from guesswork and everyday opinion by the fact that its conclusions are <u>certain</u> or at least have a vast preponderance of probability in their favor; Popper would say that scientific conclusions are never more than guesswork, hypotheses, conjectures, and that no theory ever becomes more probable. For historical reasons, according to Stove, Popper's philosophy of science received broad acceptance by the public and the scientific community. Particularly in epidemiology, where it is impossible to control many sources of extraneous influence, the possibility that a true relationship will be obscured makes it hard to refute an epidemiologic hypothesis and therefore limits the applicability of Popper's model. (See the Bibliography for other points of view.)

"Common sense":

An alternative model of scientific progress is that of "common sense", a phenomenon of increasing interest to researchers in artificial intelligence. Consider the following situation (Judea Pearl, Cognitive Systems Laboratory, UCLA, as described in M. Mitchell Waldrop, "Causality, Structure, and Common Sense," *Science* 11 September 1987; 237:1297-1299):

> You go outside in the morning and notice the grass is wet.

> The obvious inference is that it rained during the night.

However, suppose you now learn that someone left the lawn sprinkler on during the night. Suddenly your confidence in the rain goes down considerably – upon receiving a new fact, you withdraw your original conclusion.

According to presentations at the American Association for Artificial Intelligence in July 1987 (recounted in Waldrop's article), this kind of logical flip-flop ("nonmonotonic reasoning" in the artificial intelligence community) is the epitome of common sense. It is also a blatant violation of the conventional theory of logic (based on axioms, theorems, proof of theorems). But it is typical of the kind of judgment under uncertainty that characterizes both human experts and computer-based expert systems. In common sense, causes compete, evidence cooperates. The more clues we have to support a given hypothesis, the more confident we are that the hypothesis is true.

Statistical inference and causal inference

Statistical inference is not the same as causal inference, though there is a parallelism in the inferential process itself, and statistical inference is generally employed in evaluating the data for use in causal inference. In statistical inference, data from a sample of observations are used to make inferences about the population from which they are assumed to derive. A statistical model, expressed in a null hypothesis (H_0), is "tested" against data. Based on the data, the statistical model is either accepted or rejected as an adequate explanation of the data. Rejection is a stronger statement and is usually based on a more stringent criterion (a 5% significance level means that results as strong as those observed would occur by chance only 5% of the time, whereas a typical 80% level of statistical power means that a real relationship will not appear to be "significant" 20% of the time).

But excluding an explanation based on chance does not establish causality, since there are many other possible noncausal reasons for an association to exist. The association could conceivably reflect some peculiarities of the study group, problems with the measurement of disease or exposures, or the effects of some other factor that might affect both the disease AND the putative cause. In fact, the putative risk factor may have occurred AFTER (even as a result of) the disease. In causal inference, one examines the structure and results of many investigations in an attempt to assess and, if possible, eliminate all possible noncausal reasons for observed associations.

Influence of knowledge and paradigms

Since causal inference is a process of reasoning, it is conditioned by what is believed to be true and by prevailing concepts of disease. These concepts are based on knowledge of the time, as well as on ignorance and erroneous beliefs.

Consider the case of microbial agents. The Henle-Koch Postulates (1884) for implicating a bacteria as the cause of a disease held:

1. The parasite (the original term) must be present in all who have the disease;

2. The parasite can never occur in healthy persons;

3. The parasite can be isolated, cultured and capable of passing the disease to others

have been a useful model for diseases such as anthrax, tuberculosis, and tetanus. But these postulates are not adequate for many other diseases, especially viral diseases, because of (Rivers, 1937; Evans 1978):

1. Disease production may require co-factors.

2. Viruses cannot be cultured like bacteria because viruses need living cells in which to grow.

3. Pathogenic viruses can be present without clinical disease (subclinical infections, carrier states).

When pathogens are not so toxic or virulent that their presence always brings disease, then we need to consider multiple factors and a "web" of causation.

Criteria for causal inference in epidemiology

Criteria for causal inference became an issue of importance and controversy with the establishment of the first Advisory Committee to the Surgeon General on the Health Consequences of Smoking. In its 1964 report, the Committee presented a list of "epidemiologic criteria for causality" which Sir Austin Bradford Hill subsequently elaborated in his classic 1965 Presidential Address to the newly formed Section of Occupational Medicine of the Royal Society (Hill AB. The environment and disease: association or causation? *Proc Royal Soc Medicine* 1965;58:295-300). Hill's criteria are widely recognized as a basis for inferring causality.

The basic underlying questions are:

1. Is the association real or artefactual?

2. Is the association secondary to a "real" cause?

The Bradford Hill criteria

1. <u>Strength of the association</u> – The stronger an association, the less it could merely reflect the influence of some other etiologic factor(s). This criterion includes consideration of the statistical precision (minimal influence of chance) and methodologic rigor of the existing studies with respect to bias (selection, information, and confounding).

2. <u>Consistency</u> – replication of the findings by different investigators, at different times, in different places, with different methods and the ability to convincingly explain different results.

3. <u>Specificity of the association</u> – There is an inherent relationship between specificity and strength in the sense that the more accurately defined the disease and exposure, the stronger the observed relationship should be. But the fact that one agent contributes to multiple diseases is not evidence against its role in any one disease.

4. <u>Temporality</u> – the ability to establish that the putative cause in fact preceded in time the presumed effect.

5. <u>Biological gradient</u> – incremental change in disease rates in conjunction with corresponding changes in exposure. The verification of a dose-response relationship consistent with the hypothesized conceptual model.

6. <u>Plausibility</u> – we are much readier to accept the case for a relationship that is consistent with our general knowledge and beliefs. Obviously this tendency has pitfalls, but commonsense often serves us.

7. <u>Coherence</u> – how well do <u>all</u> the observations fit with the hypothesized model to form a coherent picture?

8. <u>Experiment</u> – the demonstration that under controlled conditions changing the exposure causes a change in the outcome is of great value, some would say indispensable, for inferring causality.

9. <u>Analogy</u> – we are readier to accept arguments that resemble others we accept.

Strength of the association

- Pronounced excess of disease associated with the exposure.

- The magnitude of the ratio of incidence in the exposed to incidence in the unexposed.

- How strong is "strong"? A rule-of-thumb:

Relative risk	"Meaning"
1.1-1.3	Weak
1.4-1.7	Modest
1.8-3.0	Moderate
3-8	Strong
8-16	Very strong
16-40	Dramatic
40+	Overwhelming

Strong associations are less likely to be the result of other etiologic factors than are weak associations.

> Egs., Smoking and lung cancer; smoking and CHD.

Consistency

The association has been "repeatedly observed by different persons, in different places, circumstances and times." Consistency helps to guard against associations arising out of error or artifact. But consistently observed results are not necessarily free of bias, especially across a small number of studies, and results in different populations may differ if a causal relationships is influenced by the presence or absence of modifying variables.

Specificity

The relationship between exposure and disease is specific in various ways – a specific disease is linked with a specific exposure, specific types of exposure are more effective, etc. There is an intimate relationship between specificity and strength in the sense that the more accurately defined the disease and exposure, the stronger the observed relative risk should be.

> e.g., Schildkraut and Thompson (*Am J Epidemiol* 1988; 128:456) reasoned that the familial aggregation they observed for ovarian cancer was unlikely to be due to family information bias because of the specificity of the relationship in that case-control differences in family history (a) involved malignant but not borderline disease and (b) were greater for ovarian than for other cancers.

But the fact that one agent contributes to multiple diseases is not evidence against its role in any one disease. For example, cigarette smoke causes many diseases.

Temporality

First exposure, then disease.

It is sometimes difficult to document sequence, especially if there is a long lag between the exposure and the disease, subclinical disease, exposure (e.g., a treatment) brought on by an early manifestation of the disease).

Biological gradient

The verification of a dose-response relationship consistent with the hypothesized conceptual model.
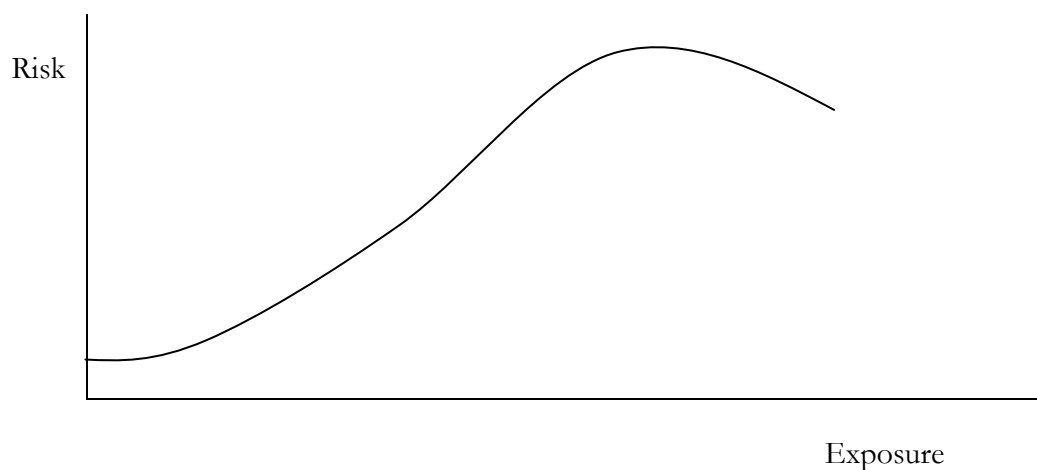
Need to consider threshold and saturation effects, characteristics of the exposure.

[See Noel Weiss, Inferring causal relationships, *Am J Epidemiol* 1981; 113:487.]

Plausibility

Does the association make sense biologically.

> E.g.s, estrogen and endometrial cancer, estrogens and breast cancer, oral contraceptives and breast cancer

Coherence

Does a causal interpretation fit with known facts of the natural history and biology of the disease, including knowledge about the distributions of the exposure and disease (by person, place, time) and the results of laboratory experiments. Do all the "pieces fit into place"?

(For an exquisite example of the evaluation of coherence, see Cornfield, Jerome, William Haenszel, E. Cuyler Hammond, Abraham M. Lilienfeld, Michael B. Shimkin, and Ernest L. Wynder. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Nat Cancer Inst* 1959;22:173–203.)

Experimental evidence

Certain types of study designs may provide more convincing evidence than other types of study designs. Intervention studies can provide the strongest support, especially when the exposure can be randomly assigned. Since it is unethical and/or impractical to assign many of the exposures that epidemiologists study. One possible alternative is to remove the exposure and see if the disease decreases, unless the causal process is regarded as irreversible.

> E.g.s, pellagra, scurvy, HDFP, LRC-CPPT, MRFIT.

Analogy

Have there been similar situations in the past? (e.g., rubella, thalidomide during pregnancy)

Except for temporality, no criterion is absolute, since causal associations can be weak, relatively nonspecific, inconsistently observed, and in conflict with prevailing biological understanding. But each criterion that is met strengthens our assurance in reaching a judgment of causality. [See also Hill's comments on tests of statistical significance.]

Several of his criteria (for example, coherence, biological gradient, specificity, and perhaps strength) may be reformulated in terms of a more general issue of consistency of observed data with a hypothesized etiological (usually biological) model. For example, a biological gradient need not be monotonic, as in the case of high doses of radiation which may lead to cell-killing and therefore a lower probability of tumor development. Similarly, specificity applies in certain situations but not others, depending upon the pathophysiologic processes hypothesized.

Search for Cause versus Decision-making

Causal inference is of fundamental importance for advancing scientific knowledge. The Popperian stance is that in an ultimate sense, every theory is tentative. Any theory can potentially be overthrown by incompatible data that cannot themselves be called into question. So in the view of many, scientific knowledge advances through concerted attempts to refute existing theories.

In considering issues in causal inference in epidemiology, though, it is useful to draw a distinction between inference aimed at establishing etiology and inference aimed at reaching a decision to act or not to act. The Popperian stance has less applicability in causal inference in support of decision-making, because of the importance of timely action. Even though individual and collective decisions are often made based on considerations other than scientific knowledge, and even without any valid causal data, causal inference is fundamental for decision-making. Moreover, judgments of causality – ultimately by governmental authorities and the public at large – are a critical basis for the resolution of controversial issues, e.g., restrictions on products such as tobacco, saccharin, coffee, oral contraceptives, handguns; pollution controls, etc. Those moved to action can cite Hill's words:

> All scientific work is incomplete - whether it be observational or experimental. All
> scientific work is liable to be upset or modified by advancing knowledge. That does not

confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.

A.B. Hill, The environment and causation, p. 300

Parallel concepts in epidemiologic inference and the legal process

One can draw an interesting analogy between the process of decision-making in epidemiology and in the legal process. In both processes, a decision about facts must be reached on the evidence available. In the absence of revealed truth (e.g., mathematical proof), both approaches emphasize integrity of the process of gathering and presenting information, adequate representation of contending views, rules of evidence, standards of certainty for various potential consequences. Both areas emphasize procedural (methodological) safeguards, since the facts in a given situation are generally established only as the findings of an adequate investigatory process. Similarly, it is important for both epidemiology and law not only that justice (i.e., proper procedures/methodology) be done but also that it <u>be seen</u> to be done. In law, pattern jury instructions provide a basis for the jury to use in weighing evidence. Similarly, epidemiology has criteria for causal inference.

The legal rules of evidence offer several parallels to the epidemiologic approach to weighing evidence and inferring causality. In both systems, reliability of the information (data) is a prime rationale. Some examples are:

- The Hearsay Rule: evidence is not admissible if based on hearsay rather than direct observation.

  Example: If the doctor testifies that the patient <u>said</u> he was driving on the wrong side of the road, that testimony is hearsay evidence and therefore not admissible. The doctor did not <u>see</u> the patient driving on the wrong side of the road.

  There are exceptions: official government sources, business records obtained in the regular course of business (<u>without an eye to a lawsuit</u>), other records routinely made are admissible in evidence.

- Dead man's statute: testimony about conversation with person who is now deceased is not admissible (because he/she cannot respond).

In both law and epidemiology, there is a relationship between the seriousness of the action and the degree of evidence required for that action. Some examples concerning searches, seizures and judgments:

- To issue a search warrant, the magistrate must find that there is a <u>reasonable suspicion</u> that the object of the search will be found.

- To issue an arrest warrant, the magistrate must find that there is <u>probable cause</u> that the person committed the crime.

- For a police officer to arrest an individual without a warrant, he must have <u>reasonable cause</u> to believe that a crime may be imminent or just committed.

- To issue an indictment, the grand jury must find that there is a <u>prima facie case</u> that the individual did commit the crime.

- For a decision against the defendant in a civil suit, the judge or jury must find a "preponderance of the evidence".

- To convict the defendant in a criminal trial, the jury must find that the evidence establishes his/her guilt "beyond a reasonable doubt".

- For a verdict of guilt based entirely on circumstantial evidence, the jury must be satisfied that <u>every reasonable hypothesis has been excluded</u> except guilt. (If there is some real evidence, the requirement is not so strict.)

(In Scotland, there is a verdict of "not proved", which certainly has parallels in epidemiologic "judgments".)

In both law and epidemiology, the facts in any individual case always factor importantly into the decision, and the decision is generally influenced by considerations of:

- How imperative is it to act?

- How imminent is a possible harm?

- How serious is the potential harm?

It is generally better to err on the side of safety (though in law that's kept implicit, never given as explicit reason).

Bibliography

Buck, Carol. Popper's philosophy for epidemiologists. *Intl J Epidemiol* 1975; 4:159-168.

Dubos, Rene. *Man adapting.* New Haven CT, Yale University, 1965.

Evans AS. Causation and disease: a chronological journey. The Thomas Parran Lecture. *Am J Epidemiol* 1978;108:249-258. See also his book of the same title, NY, Plenum, 1993.

Faust, David; Jay Ziskin. The expert witness in psychology and psychiatry. *Science* 1988; 241:31-35.

Ginzburg, Harold M. Use and misuse of epidemiologic data in the courtroom: defining the limits of inferential and particularistic evidence in mass tort litigation. *American Journal of Law & Medicine* 1986; 12 (3-4):423-439.

Hill, Austin Bradford. The environment and disease: association or causation? *Proceedings Royal Society Medicine* 1965;58:295-300. (a classic)

Jacobsen, M. Against Popperized epidemiology. *Intl J Epidemiol* 1976; 5:9-11.

Lave, Lester B.; E.P. Seskin. Epidemiology, causality and public policy. *American Scientist* 1979; 67:178-180.

Maclure, Malcolm. Popperian refutation in epidemiology. *Am J Epidemiol* 1985; 121:343-350.

Pearce, Neil; Douglas Crawford-Brown. Critical discussion in epidemiology: problems with the Popperian approach. *J Clin Epidemiol* 1989 42(3):201-208. Response by Carol Buck: 185-188.

Research Methodology: Strenghening Causal Interpretations of Nonexperimental Data (DHHS Publication No. (PHS)90-3454) - single copies available from the Agency for Health Care Policy and Research. Contains the proceedings of an AHCPR-sponsored (HS05306) conference.

Rothman and Greenland, *Modern Epidemiology*; Rothman, *Modern Epidemiology*. Ch 2.

Rothman, Kenneth, Stephen F. Lanes (eds). *Causal inference.* Chestnut Hill, MA, Epidemiology Resources Inc. W 61 C374 1988. (Includes four papers from the 1985 Society for Epidemiologic Research symposium, with solicited written criticism and rejoinders).

Schlesselman, James J. "Proof" of cause and effect in epidemiologic studies. *Preventive Medicine* 1987; 16:195-210 (from a Workshop on Guidelines to the Epidemiology of Weak Associations)

Schlesinger, George. *The intelligibility of nature.* New Jersey, Humanities Press, 1985.

Stove, David. *Popper and after.* Oxford, 1983.

Susser, Mervyn. Judgment and causal inference. Criteria in epidemiologic studies. *Am J Epidemiol* 105:1-15, 1973.

Susser, Mervyn. What is a cause and how do we know one? *Am J Epidemiol* 1991; 133:635-648.

Susser, Mervyn. The logic of Sir Karl Popper and the practice of epidemiology. *Am J Epidemiol* 1986; 124:711-718.

Weed, Douglas L. On the logic of causal inference. *Am J Epidemiol* 1986; 123:965-979 (and correspondence in 1987; 126:155-158.

Weiss, Noel S. Inferring causal relationships. *Am J Epidemiol* 1981; 113:487-