

# Causation and Causal Inference

Kenneth J. Rothman, Sander Greenland,  
Charles Poole, and Timothy L. Lash

<b>Causality 5</b>	
A Model of Sufficient Cause and Component Causes 6	
The Need for a Specific Reference Condition 7	
Application of the Sufficient-Cause Model to Epidemiology 8	
Probability, Risk, and Causes 9	
Strength of Effects 10	
Interaction among Causes 13	
Proportion of Disease due to Specific Causes 13	
Induction Period 15	
Scope of the Model 17	
Other Models of Causation 18	
<b>Philosophy of Scientific Inference 18</b>	
Inductivism 18	
Refutationism 20	
Consensus and Naturalism 21	
Bayesianism 22	
Impossibility of Scientific Proof 24	
<b>Causal Inference in Epidemiology 25</b>	
Tests of Competing Epidemiologic Theories 25	
Causal Criteria 26	

## CAUSALITY

A rudimentary understanding of cause and effect seems to be acquired by most people on their own much earlier than it could have been taught to them by someone else. Even before they can speak, many youngsters understand the relation between crying and the appearance of a parent or other adult, and the relation between that appearance and getting held, or fed. A little later, they will develop theories about what happens when a glass containing milk is dropped or turned over, and what happens when a switch on the wall is pushed from one of its resting positions to another. While theories such as these are being formulated, a more general causal theory is also being formed. The more general theory posits that some events or states of nature are causes of specific effects. Without a general theory of causation, there would be no skeleton on which to hang the substance of the many specific causal theories that one needs to survive.

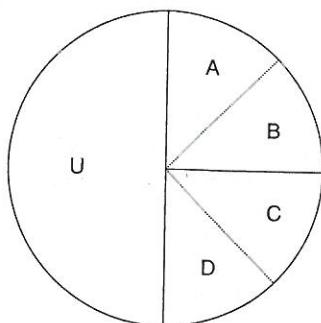
Nonetheless, the concepts of causation that are established early in life are too primitive to serve well as the basis for scientific theories. This shortcoming may be especially true in the health and social sciences, in which typical causes are neither necessary nor sufficient to bring about effects of interest. Hence, as has long been recognized in epidemiology, there is a need to develop a more refined conceptual model that can serve as a starting point in discussions of causation. In particular, such a model should address problems of multifactorial causation, confounding, interdependence of effects, direct and indirect effects, levels of causation, and systems or webs of causation (MacMahon and Pugh, 1967; Susser, 1973). This chapter describes one starting point, the sufficient-component cause model (or sufficient-cause model), which has proven useful in elucidating certain concepts in individual mechanisms of causation. Chapter 4 introduces the widely used potential-outcome or counterfactual model of causation, which is useful for relating individual-level to population-level causation, whereas Chapter 12 introduces graphical causal models (causal diagrams), which are especially useful for modeling causal systems.

Except where specified otherwise (in particular, in Chapter 27, on infectious disease), throughout the book we will assume that disease refers to a nonrecurrent event, such as death or first occurrence of a disease, and that the outcome of each individual or unit of study (e.g., a group of persons) is not affected by the exposures and outcomes of other individuals or units. Although this assumption will greatly simplify our discussion and is reasonable in many applications, it does not apply to contagious phenomena, such as transmissible behaviors and diseases. Nonetheless, all the definitions and most of the points we make (especially regarding validity) apply more generally. It is also essential to understand simpler situations before tackling the complexities created by causal interdependence of individuals or units.

### A MODEL OF SUFFICIENT CAUSE AND COMPONENT CAUSES

To begin, we need to define *cause*. One definition of the cause of a specific disease occurrence is an antecedent event, condition, or characteristic that was necessary for the occurrence of the disease at the moment it occurred, given that other conditions are fixed. In other words, a cause of a disease occurrence is an event, condition, or characteristic that preceded the disease onset and that, had the event, condition, or characteristic been different in a specified way, the disease either would not have occurred at all or would not have occurred until some later time. Under this definition, if someone walking along an icy path falls and breaks a hip, there may be a long list of causes. These causes might include the weather on the day of the incident, the fact that the path was not cleared for pedestrians, the choice of footgear for the victim, the lack of a handrail, and so forth. The constellation of causes required for this particular person to break her hip at this particular time can be depicted with the sufficient cause diagrammed in Figure 2–1. By *sufficient cause* we mean a complete causal mechanism, a minimal set of conditions and events that are sufficient for the outcome to occur. The circle in the figure comprises five segments, each of which represents a causal component that must be present or have occurred in order for the person to break her hip at that instant. The first component, labeled A, represents poor weather. The second component, labeled B, represents an uncleared path for pedestrians. The third component, labeled C, represents a poor choice of footgear. The fourth component, labeled D, represents the lack of a handrail. The final component, labeled U, represents all of the other unspecified events, conditions, and characteristics that must be present or have occurred at the instance of the fall that led to a broken hip. For etiologic effects such as the causation of disease, many and possibly all of the components of a sufficient cause may be unknown (Rothman, 1976a). We usually include one component cause, labeled U, to represent the set of unknown factors.

All of the component causes in the sufficient cause are required and must be present or have occurred at the instance of the fall for the person to break a hip. None is superfluous, which means that blocking the contribution of any component cause prevents the sufficient cause from acting. For many people, early causal thinking persists in attempts to find single causes as explanations for observed phenomena. But experience and reasoning show that the causal mechanism for any effect must consist of a constellation of components that act in concert (Mill, 1862; Mackie, 1965). In disease etiology, a sufficient cause is a set of conditions sufficient to ensure that the outcome will occur. Therefore, completing a sufficient cause is tantamount to the onset of disease. Onset here may refer to the onset of the earliest stage of the disease process or to any transition from one well-defined and readily characterized stage to the next, such as the onset of signs or symptoms.



**FIGURE 2-1** • Depiction of the constellation of component causes that constitute a sufficient cause for hip fracture for a particular person at a particular time. In the diagram, A represents poor weather, B represents an uncleared path for pedestrians, C represents a poor choice of footgear, D represents the lack of a handrail, and U represents all of the other unspecified events, conditions, and characteristics that must be present or must have occurred at the instance of the fall that led to a broken hip.

ous disease), throughout death or first occurrence group of persons) is not true. This assumption will not apply to contagious diseases and most likely. It is also essential to causal interdependence.

disease occurrence is an occurrence of the disease events, a cause of a disease disease onset and that, had the disease either would not have occurred. Under this definition, there is a long list of causes. Note that the path was not a handrail, and so forth. A hip at this particular time. By *sufficient cause* we mean that there are sufficient for each of which represents a component that leads to break her hip at that point. The final component, labeled C, represents a poor condition of a handrail. The final components, and characteristics of a broken hip. For etiologic components of a sufficient cause, labeled U, to

must be present or have been present, which means that the cause from acting. Causes as explanations causal mechanism for any (Hume, 1862; Mackie, 1965). I am sure that the outcome onset of disease. Onset may transition from one signs or symptoms.

tion of component fracture for a particular represents poor weather, C represents a poor handrail, and U conditions, and have occurred at the

Consider again the role of the handrail in causing hip fracture. The absence of such a handrail may play a causal role in some sufficient causes but not in others, depending on circumstances such as the weather, the level of inebriation of the pedestrian, and countless other factors. Our definition links the lack of a handrail with this one broken hip and does not imply that the lack of this handrail by itself was sufficient for that hip fracture to occur. With this definition of cause, no specific event, condition, or characteristic is sufficient by itself to produce disease. The definition does not describe a complete causal mechanism, but only a component of it. To say that the absence of a handrail is a component cause of a broken hip does not, however, imply that every person walking down the path will break a hip. Nor does it imply that if a handrail is installed with properties sufficient to prevent that broken hip, that no one will break a hip on that same path. There may be other sufficient causes by which a person could suffer a hip fracture. Each such sufficient cause would be depicted by its own diagram similar to Figure 2-1. The first of these sufficient causes to be completed by simultaneous accumulation of all of its component causes will be the one that depicts the mechanism by which the hip fracture occurs for a particular person. If no sufficient cause is completed while a person passes along the path, then no hip fracture will occur over the course of that walk.

As noted above, a characteristic of the naive concept of causation is the assumption of a one-to-one correspondence between the observed cause and effect. Under this view, each cause is seen as "necessary" and "sufficient" in itself to produce the effect, particularly when the cause is an observable action or event that takes place near in time to the effect. Thus, the flick of a switch appears to be the singular cause that makes an electric light go on. There are less evident causes, however, that also operate to produce the effect: a working bulb in the light fixture, intact wiring from the switch to the bulb, and voltage to produce a current when the circuit is closed. To achieve the effect of turning on the light, each of these components is as important as moving the switch, because changing any of these components of the causal constellation will prevent the effect. The term *necessary cause* is therefore reserved for a particular type of component cause under the sufficient-cause model. If any of the component causes appears in every sufficient cause, then that component cause is called a "necessary" component cause. For the disease to occur, any and all necessary component causes must be present or must have occurred. For example, one could label a component cause with the requirement that one must have a hip to suffer a hip fracture. Every sufficient cause that leads to hip fracture must have that component cause present, because in order to fracture a hip, one must have a hip to fracture.

The concept of complementary component causes will be useful in applications to epidemiology that follow. For each component cause in a sufficient cause, the set of the other component causes in that sufficient cause comprises the complementary component causes. For example, in Figure 2-1, component cause A (poor weather) has as its complementary component causes the components labeled B, C, D, and U. Component cause B (an uncleared path for pedestrians) has as its complementary component causes the components labeled A, C, D, and U.

### THE NEED FOR A SPECIFIC REFERENCE CONDITION

Component causes must be defined with respect to a clearly specified alternative or reference condition (often called a *referent*). Consider again the lack of a handrail along the path. To say that this condition is a component cause of the broken hip, we have to specify an alternative condition against which to contrast the cause. The mere presence of a handrail would not suffice. After all, the hip fracture might still have occurred in the presence of a handrail, if the handrail was too short or if it was old and made of rotten wood. We might need to specify the presence of a handrail sufficiently tall and sturdy to break the fall for the absence of that handrail to be a component cause of the broken hip.

To see the necessity of specifying the alternative event, condition, or characteristic as well as the causal one, consider an example of a man who took high doses of ibuprofen for several years and developed a gastric ulcer. Did the man's use of ibuprofen cause his ulcer? One might at first assume that the natural contrast would be with what would have happened had he taken nothing instead of ibuprofen. Given a strong reason to take the ibuprofen, however, that alternative may not make sense. If the specified alternative to taking ibuprofen is to take acetaminophen, a different drug that might have been indicated for his problem, and if he would not have developed the ulcer had he used acetaminophen, then we can say that using ibuprofen caused the ulcer. But ibuprofen did not cause

his ulcer if the specified alternative is taking aspirin and, had he taken aspirin, he still would have developed the ulcer. The need to specify the alternative to a preventive is illustrated by a newspaper headline that read: "Rare Meat Cuts Colon Cancer Risk." Was this a story of an epidemiologic study comparing the colon cancer rate of a group of people who ate rare red meat with the rate in a group of vegetarians? No, the study compared persons who ate rare red meat with persons who ate highly cooked red meat. The same exposure, regular consumption of rare red meat, might have a preventive effect when contrasted against highly cooked red meat and a causative effect or no effect in contrast to a vegetarian diet. An event, condition, or characteristic is not a cause by itself as an intrinsic property it possesses in isolation, but as part of a causal contrast with an alternative event, condition, or characteristic (Lewis, 1973; Rubin, 1974; Greenland et al., 1999a; Maldonado and Greenland, 2002; see Chapter 4).

### APPLICATION OF THE SUFFICIENT-CAUSE MODEL TO EPIDEMIOLOGY

The preceding introduction to concepts of sufficient causes and component causes provides the lexicon for application of the model to epidemiology. For example, tobacco smoking is a cause of lung cancer, but by itself it is not a sufficient cause, as demonstrated by the fact that most smokers do not get lung cancer. First, the term *smoking* is too imprecise to be useful beyond casual description. One must specify the type of smoke (e.g., cigarette, cigar, pipe, or environmental), whether it is filtered or unfiltered, the manner and frequency of inhalation, the age at initiation of smoking, and the duration of smoking. And, however smoking is defined, its alternative needs to be defined as well. Is it smoking nothing at all, smoking less, smoking something else? Equally important, even if smoking and its alternative are both defined explicitly, smoking will not cause cancer in everyone. So who is susceptible to this smoking effect? Or, to put it in other terms, what are the other components of the causal constellation that act with smoking to produce lung cancer in this contrast?

Figure 2-2 provides a schematic diagram of three sufficient causes that could be completed during the follow-up of an individual. The three conditions or events—A, B, and E—have been defined as binary variables, so they can only take on values of 0 or 1. With the coding of A used in the figure, its reference level,  $A = 0$ , is sometimes causative, but its index level,  $A = 1$ , is never causative. This situation arises because two sufficient causes contain a component cause labeled " $A = 0$ ," but no sufficient cause contains a component cause labeled " $A = 1$ ." An example of a condition or event of this sort might be  $A = 1$  for taking a daily multivitamin supplement and  $A = 0$  for taking no vitamin supplement. With the coding of B and E used in the example depicted by Figure 2-2, their index levels,  $B = 1$  and  $E = 1$ , are sometimes causative, but their reference levels,  $B = 0$  and  $E = 0$ , are never causative. For each variable, the index and reference levels may represent only two alternative states or events out of many possibilities. Thus, the coding of B might be  $B = 1$  for smoking 20 cigarettes per day for 40 years and  $B = 0$  for smoking 20 cigarettes per day for 20 years, followed by 20 years of not smoking. E might be coded  $E = 1$  for living in an urban neighborhood with low average income and high income inequality, and  $E = 0$  for living in an urban neighborhood with high average income and low income inequality.

$A = 0$ ,  $B = 1$ , and  $E = 1$  are individual component causes of the sufficient causes in Figure 2-2.  $U_1$ ,  $U_2$ , and  $U_3$  represent sets of component causes.  $U_1$ , for example, is the set of all components other than  $A = 0$  and  $B = 1$  required to complete the first sufficient cause in Figure 2-2. If we decided not to specify  $B = 1$ , then  $B = 1$  would become part of the set of components that are causally complementary to  $A = 0$ ; in other words,  $B = 1$  would then be absorbed into  $U_1$ .

Each of the three sufficient causes represented in Figure 2-2 is minimally sufficient to produce the disease in the individual. That is, only one of these mechanisms needs to be completed for

*(can be more than one "sufficient cause")*

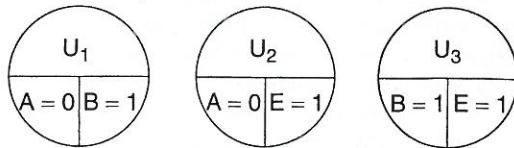


FIGURE 2-2 • Three classes of sufficient causes of a disease (sufficient causes I, II, and III from left to right).

still would have been reported by a newspaper if an epidemiologic study with the rate in which persons who eat meat, might have a causative effect or not be a cause by itself with an alternative (Maldonado 1999a; Maldonado

uses provides the smoking is a cause of it most smokers do a causal description. (ital), whether it is a cause of smoking, needs to be defined. Equally important, if cause cancer in terms, what are the lung cancer in this

ould be completed and E—have been coding of A used alone, A = 1, is never a sufficient cause labeled. An example of a supplement and example depicted but their reference levels may coding of B might be 20 cigarettes per day for living in an area where U = 0 for living in an

ses in Figure 2–2. of all components Figure 2–2. If we components that are put into U<sub>1</sub>. Sufficient to produce be completed for Case 1)

uses of sufficient causes I, II, and III

disease to occur (sufficiency), and there is no superfluous component cause in any mechanism (minimality)—each component is a required part of that specific causal mechanism. A specific component cause may play a role in one, several, or all of the causal mechanisms. As noted earlier, a component cause that appears in all sufficient causes is called a *necessary* cause of the outcome. As an example, infection with HIV is a component of every sufficient cause of acquired immune deficiency syndrome (AIDS) and hence is a necessary cause of AIDS. It has been suggested that such causes be called “universally necessary,” in recognition that every component of a sufficient cause is necessary for that sufficient cause (mechanism) to operate (Poole 2001a).

Figure 2–2 does not depict aspects of the causal process such as sequence or timing of action of the component causes, dose, or other complexities. These can be specified in the description of the contrast of index and reference conditions that defines each component cause. Thus, if the outcome is lung cancer and the factor B represents cigarette smoking, it might be defined more explicitly as smoking at least 20 cigarettes a day of unfiltered cigarettes for at least 40 years beginning at age 20 years or earlier (B = 1), or smoking 20 cigarettes a day of unfiltered cigarettes, beginning at age 20 years or earlier, and then smoking no cigarettes for the next 20 years (B = 0).

In specifying a component cause, the two sides of the causal contrast of which it is composed should be defined with an eye to realistic choices or options. If prescribing a placebo is not a realistic therapeutic option, a causal contrast between a new treatment and a placebo in a clinical trial may be questioned for its dubious relevance to medical practice. In a similar fashion, before saying that oral contraceptives increase the risk of death over 10 years (e.g., through myocardial infarction or stroke), we must consider the alternative to taking oral contraceptives. If it involves getting pregnant, then the risk of death attendant to childbirth might be greater than the risk from oral contraceptives, making oral contraceptives a preventive rather than a cause. If the alternative is an equally effective contraceptive without serious side effects, then oral contraceptives may be described as a cause of death.

To understand prevention in the sufficient-component cause framework, we posit that the alternative condition (in which a component cause is absent) prevents the outcome relative to the presence of the component cause. Thus, a preventive effect of a factor is represented by specifying its causative alternative as a component cause. An example is the presence of A = 0 as a component cause in the first two sufficient causes shown in Figure 2–2. Another example would be to define a variable, F (not depicted in Fig. 2–2), as “vaccination (F = 1) or no vaccination (F = 0)”. Prevention of the disease by getting vaccinated (F = 1) would be expressed in the sufficient-component cause model as causation of the disease by not getting vaccinated (F = 0). This depiction is unproblematic because, once both sides of a causal contrast have been specified, causation and prevention are merely two sides of the same coin.

Sheps (1958) once asked, “Shall we count the living or the dead?” Death is an event, but survival is not. Hence, to use the sufficient-component cause model, we must count the dead. This model restriction can have substantive implications. For instance, some measures and formulas approximate others only when the outcome is rare. When survival is rare, death is common. In that case, use of the sufficient-component cause model to inform the analysis will prevent us from taking advantage of the rare-outcome approximations.

Similarly, etiologies of adverse health outcomes that are conditions or states, but not events, must be depicted under the sufficient-cause model by reversing the coding of the outcome. Consider spina bifida, which is the failure of the neural tube to close fully during gestation. There is no point in time at which spina bifida may be said to have occurred. It would be awkward to define the “incidence time” of spina bifida as the gestational age at which complete neural tube closure ordinarily occurs. The sufficient-component cause model would be better suited in this case to defining the event of complete closure (no spina bifida) as the outcome and to view conditions, events, and characteristics that prevent this beneficial event as the causes of the adverse condition of spina bifida.

### PROBABILITY, RISK, AND CAUSES

In everyday language, “risk” is often used as a synonym for probability. It is also commonly used as a synonym for “hazard,” as in, “Living near a nuclear power plant is a risk you should avoid.” Unfortunately, in epidemiologic parlance, even in the scholarly literature, “risk” is frequently used for many distinct concepts: rate, rate ratio, risk ratio, incidence odds, prevalence, etc. The more

specific, and therefore more useful, definition of *risk* is “probability of an event during a specified period of time.”

The term *probability* has multiple meanings. One is that it is the relative frequency of an event. Another is that *probability* is the tendency, or propensity, of an entity to produce an event. A third meaning is that probability measures someone’s degree of certainty that an event will occur. When one says “the probability of death in vehicular accidents when traveling >120 km/h is high,” one means that the proportion of accidents that end with deaths is higher when they involve vehicles traveling >120 km/h than when they involve vehicles traveling at lower speeds (frequency usage), that high-speed accidents have a greater tendency than lower-speed accidents to result in deaths (propensity usage), or that the speaker is more certain that a death will occur in a high-speed accident than in a lower-speed accident (certainty usage).

The frequency usage of “*probability*” and “*risk*,” unlike the propensity and certainty usages, admits no meaning to the notion of “*risk*” for an individual beyond the relative frequency of 100% if the event occurs and 0% if it does not. This restriction of individual risks to 0 or 1 can only be relaxed to allow values in between by reinterpreting such statements as the frequency with which the outcome would be seen upon random sampling from a very large population of individuals deemed to be “like” the individual in some way (e.g., of the same age, sex, and smoking history). If one accepts this interpretation, whether any actual sampling has been conducted or not, the notion of individual risk is replaced by the notion of the frequency of the event in question in the large population from which the individual was sampled. With this view of risk, a risk will change according to how we group individuals together to evaluate frequencies. Subjective judgment will inevitably enter into the picture in deciding which characteristics to use for grouping. For instance, should tomato consumption be taken into account in defining the class of men who are “like” a given man for purposes of determining his risk of a diagnosis of prostate cancer between his 60th and 70th birthdays? If so, which study or meta-analysis should be used to factor in this piece of information?

Unless we have found a set of conditions and events in which the disease does not occur at all, it is always a reasonable working hypothesis that, no matter how much is known about the etiology of a disease, some causal components remain unknown. We may be inclined to assign an equal risk to all individuals whose status for some components is known and identical. We may say, for example, that men who are heavy cigarette smokers have approximately a 10% lifetime risk of developing lung cancer. Some interpret this statement to mean that all men would be subject to a 10% probability of lung cancer if they were to become heavy smokers, as if the occurrence of lung cancer, aside from smoking, were purely a matter of chance. This view is untenable. A probability may be 10% conditional on one piece of information and higher or lower than 10% if we condition on other relevant information as well. For instance, men who are heavy cigarette smokers and who worked for many years in occupations with historically high levels of exposure to airborne asbestos fibers would be said to have a lifetime lung cancer risk appreciably higher than 10%.

Regardless of whether we interpret probability as relative frequency or degree of certainty, the assignment of equal risks merely reflects the particular grouping. In our ignorance, the best we can do in assessing risk is to classify people according to measured risk indicators and then assign the average risk observed within a class to persons within the class. As knowledge or specification of additional risk indicators expands, the risk estimates assigned to people will depart from average according to the presence or absence of other factors that predict the outcome.

### STRENGTH OF EFFECTS

The causal model exemplified by Figure 2–2 can facilitate an understanding of some key concepts such as *strength of effect* and *interaction*. As an illustration of strength of effect, Table 2–1 displays the frequency of the eight possible patterns for exposure to A, B, and E in two hypothetical populations. Now the pie charts in Figure 2–2 depict classes of mechanisms. The first one, for instance, represents all sufficient causes that, no matter what other component causes they may contain, have in common the fact that they contain A = 0 and B = 1. The constituents of U<sub>1</sub> may, and ordinarily would, differ from individual to individual. For simplification, we shall suppose, rather unrealistically, that U<sub>1</sub>, U<sub>2</sub>, and U<sub>3</sub> are always present or have always occurred for everyone and Figure 2–2 represents all the sufficient causes.

nt during a specified

equency of an event. ice an event. A third ent will occur. When 0 km/h is high,” one key involve vehicles ls (frequency usage), ts to result in deaths i high-speed accident

nd certainty usages, e frequency of 100% o 0 or 1 can only be equency with which lation of individuals d smoking history). onducted or not, the nt in question in the sk, a risk will change ective judgment will ouping. For instance, en who are “like” a er between his 60th actor in this piece of

does not occur at all, n about the etiology d to assign an equal cal. We may say, for 10% lifetime risk of would be subject to a e occurrence of lung nable. A probability 10% if we condition tte smokers and who : to airborne asbestos n 10%. gree of certainty, the nce, the best we can s and then assign the ge or specification of depart from average

f some key concepts t, Table 2–1 displays o hypothetical popu- rst one, for instance, ey may contain, have may, and ordinarily use, rather unrealisti- yone and Figure 2–2

**TABLE 2–1**

**Exposure Frequencies and Individual Risks in Two Hypothetical Populations According to the Possible Combinations of the Three Specified Component Causes in Fig. 2–1**

Exposures				Risk	Frequency of Exposure Pattern	
A	B	E	Sufficient Cause Completed		Population 1	Population 2
1	1	1	III	1	900	100
1	1	0	None	0	900	100
1	0	1	None	0	100	900
1	0	0	None	0	100	900
0	1	1	I, II, or III	1	100	900
0	1	0	I	1	100	900
0	0	1	II	1	900	100
0	0	0	none	0	900	100

Under these assumptions, the response of each individual to the exposure pattern in a given row can be found in the response column. The response here is the risk of developing a disease over a specified time period that is the same for all individuals. For simplification, a deterministic model of risk is employed, such that individual risks can equal only the value 0 or 1, and no values in between. A stochastic model of individual risk would relax this restriction and allow individual risks to lie between 0 and 1.

The proportion getting disease, or incidence proportion, in any subpopulation in Table 2–1 can be found by summing the number of persons at each exposure pattern with an individual risk of 1 and dividing this total by the subpopulation size. For example, if exposure A is not considered (e.g., if it were not measured), the pattern of incidence proportions in population 1 would be those in Table 2–2.

As an example of how the proportions in Table 2–2 were calculated, let us review how the incidence proportion among persons in population 1 with B = 1 and E = 0 was calculated: There were 900 persons with A = 1, B = 1, and E = 0, none of whom became cases because there are no sufficient causes that can culminate in the occurrence of the disease over the study period in persons with this combination of exposure conditions. (There are two sufficient causes that contain B = 1 as a component cause, but one of them contains the component cause A = 0 and the other contains the component cause E = 1. The presence of A = 1 or E = 0 blocks these etiologic mechanisms.) There were 100 persons with A = 0, B = 1, and E = 0, all of whom became cases because they all had U<sub>1</sub>, the set of causal complements for the class of sufficient causes containing A = 0 and

**TABLE 2–2**

**Incidence Proportions (IP) for Combinations of Component Causes B and E in Hypothetical Population 1, Assuming That Component Cause A Is Unmeasured**

	B = 1, E = 1	B = 1, E = 0	B = 0, E = 1	B = 0, E = 0
Cases	1,000	100	900	0
Total	1,000	1,000	1,000	1,000
IP	1.00	0.10	0.90	0.00

TABLE 2-3

**Incidence Proportions (IP) for Combinations of Component Causes B and E in Hypothetical Population 2, Assuming That Component Cause A Is Unmeasured**

	B = 1, E = 1	B = 1, E = 0	B = 0, E = 1	B = 0, E = 0
Cases	1,000	900	100	0
Total	1,000	1,000	1,000	1,000
IP	1.00	0.90	0.10	0.00

$B = 1$ . Thus, among all 1,000 persons with  $B = 1$  and  $E = 0$ , there were 100 cases, for an incidence proportion of 0.10.

If we were to measure strength of effect by the difference of the incidence proportions, it is evident from Table 2-2 that for population 1,  $E = 1$  has a much stronger effect than  $B = 1$ , because  $E = 1$  increases the incidence proportion by 0.9 (in both levels of  $B$ ), whereas  $B = 1$  increases the incidence proportion by only 0.1 (in both levels of  $E$ ). Table 2-3 shows the analogous results for population 2. Although the members of this population have exactly the same causal mechanisms operating within them as do the members of population 1, the relative strengths of causative factors  $E = 1$  and  $B = 1$  are reversed, again using the incidence proportion difference as the measure of strength.  $B = 1$  now has a much stronger effect on the incidence proportion than  $E = 1$ , despite the fact that  $A$ ,  $B$ , and  $E$  have no association with one another in either population, and their index levels ( $A = 1$ ,  $B = 1$  and  $E = 1$ ) and reference levels ( $A = 0$ ,  $B = 0$ , and  $E = 0$ ) are each present or have occurred in exactly half of each population.

The overall difference of incidence proportions contrasting  $E = 1$  with  $E = 0$  is  $(1,900/2,000) - (100/2,000) = 0.9$  in population 1 and  $(1,100/2,000) - (900/2,000) = 0.1$  in population 2. The key difference between populations 1 and 2 is the difference in the prevalence of the conditions under which  $E = 1$  acts to increase risk: that is, the presence of  $A = 0$  or  $B = 1$ , but not both. (When  $A = 0$  and  $B = 1$ ,  $E = 1$  completes all three sufficient causes in Figure 2-2; it thus does not increase anyone's risk, although it may well shorten the time to the outcome.) The prevalence of the condition, "A = 0 or B = 1 but not both" is  $1,800/2,000 = 90\%$  in both levels of  $E$  in population 1. In population 2, this prevalence is only  $200/2,000 = 10\%$  in both levels of  $E$ . This difference in the prevalence of the conditions sufficient for  $E = 1$  to increase risk explains the difference in the strength of the effect of  $E = 1$  as measured by the difference in incidence proportions.

As noted above, the set of all other component causes in all sufficient causes in which a causal factor participates is called the *causal complement* of the factor. Thus,  $A = 0$ ,  $B = 1$ ,  $U_2$ , and  $U_3$  make up the causal complement of  $E = 1$  in the above example. This example shows that the strength of a factor's effect on the occurrence of a disease in a population, measured as the absolute difference in incidence proportions, depends on the prevalence of its causal complement. This dependence has nothing to do with the etiologic mechanism of the component's action, because the component is an equal partner in each mechanism in which it appears. Nevertheless, a factor will appear to have a strong effect, as measured by the difference of proportions getting disease, if its causal complement is common. Conversely, a factor with a rare causal complement will appear to have a weak effect.

If strength of effect is measured by the ratio of proportions getting disease, as opposed to the difference, then strength depends on more than a factor's causal complement. In particular, it depends additionally on how common or rare the components are of sufficient causes in which the specified causal factor does *not* play a role. In this example, given the ubiquity of  $U_1$ , the effect of  $E = 1$  measured in ratio terms depends on the prevalence of  $E = 1$ 's causal complement and on the prevalence of the conjunction of  $A = 0$  and  $B = 1$ . If many people have both  $A = 0$  and  $B = 1$ , the "baseline" incidence proportion (i.e., the proportion of not- $E$  or "unexposed" persons getting disease) will be high and the proportion getting disease due to  $E$  will be comparatively low. If few

onent  
ng That

**B = 0, E = 0**

0
1,000
0.00

cases, for an incidence

lence proportions, it is  
ect than  $B = 1$ , because  
as  $B = 1$  increases the  
e analogous results for  
me causal mechanisms  
ths of causative factors  
ence as the measure of  
on than  $E = 1$ , despite  
ulation, and their index  
 $\beta = 0$ ) are each present

$= 0$  is  $(1,900/2,000) - 1$  in population 2. The  
ence of the conditions  
 $r B = 1$ , but not both.  
re 2–2; it thus does not  
) The prevalence of the  
ls of  $E$  in population 1.  
f  $E$ . This difference in  
is the difference in the  
roportions.

uses in which a causal  
 $: 0, B = 1, U_2$ , and  $U_3$   
shows that the strength  
the absolute difference  
t. This dependence has  
ause the component is  
tor will appear to have  
>, if its causal comple-  
appear to have a weak

disease, as opposed to  
ement. In particular, it  
nt causes in which the  
ity of  $U_1$ , the effect of  
omplement and on the  
oth  $A = 0$  and  $B = 1$ ,  
osed” persons getting  
paratively low. If few

people have both  $A = 0$  and  $B = 1$ , the baseline incidence proportion will be low and the proportion getting disease due to  $E = 1$  will be comparatively high. Thus, strength of effect measured by the incidence proportion ratio depends on more conditions than does strength of effect measured by the incidence proportion difference.

Regardless of how strength of a causal factor's effect is measured, the public health significance of that effect does not imply a corresponding degree of etiologic significance. Each component cause in a given sufficient cause has the same etiologic significance. Given a specific causal mechanism, any of the component causes can have strong or weak effects using either the difference or ratio measure. The actual identities of the components of a sufficient cause are part of the mechanics of causation, whereas the strength of a factor's effect depends on the time-specific distribution of its causal complement (if strength is measured in absolute terms) plus the distribution of the components of all sufficient causes in which the factor does not play a role (if strength is measured in relative terms). Over a span of time, the strength of the effect of a given factor on disease occurrence may change because the prevalence of its causal complement in various mechanisms may also change, even if the causal mechanisms in which the factor and its cofactors act remain unchanged.

### INTERACTION AMONG CAUSES

Two component causes acting in the same sufficient cause may be defined as *interacting causally* to produce disease. This definition leaves open many possible mechanisms for the interaction, including those in which two components interact in a direct physical fashion (e.g., two drugs that react to form a toxic by-product) and those in which one component (the *initiator* of the pair) alters a substrate so that the other component (the *promoter* of the pair) can act. Nonetheless, it excludes any situation in which one component  $E$  is merely a cause of another component  $F$ , with no effect of  $E$  on disease except through the component  $F$  it causes.

Acting in the same sufficient cause is not the same as one component cause acting to produce a second component cause, and then the second component going on to produce the disease (Robins and Greenland 1992, Kaufman et al., 2004). As an example of the distinction, if cigarette smoking (vs. never smoking) is a component cause of atherosclerosis, and atherosclerosis (vs. no atherosclerosis) causes myocardial infarction, both smoking and atherosclerosis would be component causes (cofactors) in certain sufficient causes of myocardial infarction. They would not necessarily appear in the same sufficient cause. Rather, for a sufficient cause involving atherosclerosis as a component cause, there would be another sufficient cause in which the atherosclerosis component cause was replaced by all the component causes that brought about the atherosclerosis, including smoking. Thus, a sequential causal relation between smoking and atherosclerosis would not be enough for them to interact synergistically in the etiology of myocardial infarction, in the sufficient-cause sense. Instead, the causal sequence means that smoking can act indirectly, through atherosclerosis, to bring about myocardial infarction.

Now suppose that, perhaps in addition to the above mechanism, smoking reduces clotting time and thus causes thrombi that block the coronary arteries if they are narrowed by atherosclerosis. This mechanism would be represented by a sufficient cause containing both smoking and atherosclerosis as components and thus would constitute a synergistic interaction between smoking and atherosclerosis in causing myocardial infarction. The presence of this sufficient cause would not, however, tell us whether smoking also contributed to the myocardial infarction by causing the atherosclerosis. Thus, the basic sufficient-cause model does not alert us to indirect effects (effects of some component causes mediated by other component causes in the model). Chapters 4 and 12 introduce potential-outcome and graphical models better suited to displaying indirect effects and more general sequential mechanisms, whereas Chapter 5 discusses in detail interaction as defined in the potential-outcome framework and its relation to interaction as defined in the sufficient-cause model.

### PROPORTION OF DISEASE DUE TO SPECIFIC CAUSES

In Figure 2–2, assuming that the three sufficient causes in the diagram are the only ones operating, what fraction of disease is caused by  $E = 1$ ?  $E = 1$  is a component cause of disease in two of the sufficient-cause mechanisms, II and III, so all disease arising through either of these two mechanisms is attributable to  $E = 1$ . Note that in persons with the exposure pattern  $A = 0, B = 1, E = 1$ , all three

sufficient causes would be completed. The first of the three mechanisms to be completed would be the one that actually produces a given case. If the first one completed is mechanism II or III, the case would be causally attributable to  $E = 1$ . If mechanism I is the first one to be completed, however,  $E = 1$  would not be part of the sufficient cause producing that case. Without knowing the completion times of the three mechanisms, among persons with the exposure pattern  $A = 0, B = 1, E = 1$  we cannot tell how many of the 100 cases in population 1 or the 900 cases in population 2 are etiologically attributable to  $E = 1$ .

Each of the cases that is etiologically attributable to  $E = 1$  can also be attributed to the other component causes in the causal mechanisms in which  $E = 1$  acts. Each component cause interacts with its complementary factors to produce disease, so each case of disease can be attributed to every component cause in the completed sufficient cause. Note, though, that the attributable fractions added across component causes of the same disease do not sum to 1, although there is a mistaken tendency to think that they do. To illustrate the mistake in this tendency, note that a necessary component cause appears in every completed sufficient cause of disease, and so by itself has an attributable fraction of 1, without counting the attributable fractions for other component causes. Because every case of disease can be attributed to every component cause in its causal mechanism, attributable fractions for different component causes will generally sum to more than 1, and there is no upper limit for this sum.

A recent debate regarding the proportion of risk factors for coronary heart disease attributable to particular component causes illustrates the type of errors in inference that can arise when the sum is thought to be restricted to 1. The debate centers around whether the proportion of coronary heart disease attributable to high blood cholesterol, high blood pressure, and cigarette smoking equals 75% or "only 50%" (Magnus and Beaglehole, 2001). If the former, then some have argued that the search for additional causes would be of limited utility (Beaglehole and Magnus, 2002), because only 25% of cases "remain to be explained." By assuming that the proportion explained by yet unknown component causes cannot exceed 25%, those who support this contention fail to recognize that cases caused by a sufficient cause that contains any subset of the three named causes might also contain unknown component causes. Cases stemming from sufficient causes with this overlapping set of component causes could be prevented by interventions targeting the three named causes, or by interventions targeting the yet unknown causes when they become known. The latter interventions could reduce the disease burden by much more than 25%.

As another example, in a cohort of cigarette smokers exposed to arsenic by working in a smelter, an estimated 75% of the lung cancer rate was attributable to their work environment and an estimated 65% was attributable to their smoking (Pinto et al., 1978; Hertz-Pannier et al., 1992). There is no problem with such figures, which merely reflect the multifactorial etiology of disease. So, too, with coronary heart disease; if 75% of that disease is attributable to high blood cholesterol, high blood pressure, and cigarette smoking, 100% of it can still be attributable to other causes, known, suspected, and yet to be discovered. Some of these causes will participate in the same causal mechanisms as high blood cholesterol, high blood pressure, and cigarette smoking. Beaglehole and Magnus were correct in thinking that if the three specified component causes combine to explain 75% of cardiovascular disease (CVD) and we somehow eliminated them, there would be only 25% of CVD cases remaining. But until that 75% is eliminated, any newly discovered component could cause up to 100% of the CVD we currently have.

The notion that interventions targeting high blood cholesterol, high blood pressure, and cigarette smoking could eliminate 75% of coronary heart disease is unrealistic given currently available intervention strategies. Although progress can be made to reduce the effect of these risk factors, it is unlikely that any of them could be completely eradicated from any large population in the near term. Estimates of the public health effect of eliminating diseases themselves as causes of death (Murray et al., 2002) are even further removed from reality, because they fail to account for all the effects of interventions required to achieve the disease elimination, including unanticipated side effects (Greenland, 2002a, 2005a).

The debate about coronary heart disease attribution to component causes is reminiscent of an earlier debate regarding causes of cancer. In their widely cited work, *The Causes of Cancer*, Doll and Peto (1981, Table 20) created a table giving their estimates of the fraction of all cancers caused by various agents. The fractions summed to nearly 100%. Although the authors acknowledged that any case could be caused by more than one agent (which means that, given enough agents, the attributable

part - what is attributable risk?

o be completed would be mechanism II or III, t one to be completed, ∴ Without knowing the ure pattern A = 0, B = 00 cases in population

attributed to the other component cause interacts n be attributed to every > attributable fractions ligh there is a mistaken note that a necessary und so by itself has an er component causes. its causal mechanism, more than 1, and there

art disease attributable at can arise when the proportion of coronary and cigarette smoking hen some have argued e and Magnus, 2002), > proportion explained this contention fail to he three named causes cien causes with this geting the three named ome known. The latter

y working in a smelter, ment and an estimated et al., 1992). There is gy of disease. So, too, blood cholesterol, high > other causes, known, te in the same causal oking. Beaglehole and es combine to explain re would be only 25% ered component could

pressure, and cigarette en currently available of these risk factors, it population in the near /es as causes of death l to account for all the ng unanticipated side

s is reminiscent of an es of Cancer, Doll and all cancers caused by cknowledged that any agents, the attributable

fractions would sum to far more than 100%), they referred to this situation as a “difficulty” and an “anomaly” that they chose to ignore. Subsequently, one of the authors acknowledged that the attributable fraction could sum to greater than 100% (Peto, 1985). It is neither a difficulty nor an anomaly nor something we can safely ignore, but simply a consequence of the fact that no event has a single agent as the cause. The fraction of disease that can be attributed to known causes will grow without bound as more causes are discovered. Only the fraction of disease attributable to a single component cause cannot exceed 100%.

In a similar vein, much publicity attended the pronouncement in 1960 that as much as 90% of cancer is environmentally caused (Higginson, 1960). Here, “environment” was thought of as representing all nongenetic component causes, and thus included not only the physical environment, but also the social environment and all individual human behavior that is not genetically determined. Hence, environmental component causes must be present to some extent in every sufficient cause of a disease. Thus, Higginson’s estimate of 90% was an underestimate.

One can also show that 100% of any disease is inherited, even when environmental factors are component causes. MacMahon (1968) cited the example given by Hogben (1933) of yellow shanks, a trait occurring in certain genetic strains of fowl fed on yellow corn. Both a particular set of genes and a yellow-corn diet are necessary to produce yellow shanks. A farmer with several strains of fowl who feeds them all only yellow corn would consider yellow shanks to be a genetic condition, because only one strain would get yellow shanks, despite all strains getting the same diet. A different farmer who owned only the strain liable to get yellow shanks but who fed some of the birds yellow corn and others white corn would consider yellow shanks to be an environmentally determined condition because it depends on diet. In humans, the mental retardation caused by phenylketonuria is considered by many to be purely genetic. This retardation can, however, be successfully prevented by dietary intervention, which demonstrates the presence of an environmental cause. In reality, yellow shanks, phenylketonuria, and other diseases and conditions are determined by an interaction of genes and environment. It makes no sense to allocate a portion of the causation to either genes or environment separately when both may act together in sufficient causes.

Nonetheless, many researchers have compared disease occurrence in identical and nonidentical twins to estimate the fraction of disease that is inherited. These twin-study and other heritability indices assess only the relative role of environmental and genetic causes of disease in a particular setting. For example, some genetic causes may be necessary components of every causal mechanism. If everyone in a population has an identical set of the genes that cause disease, however, their effect is not included in heritability indices, despite the fact that the genes are causes of the disease. The two farmers in the preceding example would offer very different values for the heritability of yellow shanks, despite the fact that the condition is always 100% dependent on having certain genes.

Every case of every disease has some environmental and some genetic component causes, and therefore every case can be attributed both to genes and to environment. No paradox exists as long as it is understood that the fractions of disease attributable to genes and to environment overlap with one another. Thus, debates over what proportion of all occurrences of a disease are genetic and what proportion are environmental, inasmuch as these debates assume that the shares must add up to 100%, are fallacious and distracting from more worthwhile pursuits.

On an even more general level, the question of whether a given disease does or does not have a “multifactorial etiology” can be answered once and for all in the affirmative. All diseases have multifactorial etiologies. It is therefore completely unremarkable for a given disease to have such an etiology, and no time or money should be spent on research trying to answer the question of whether a particular disease does or does not have a multifactorial etiology. They all do. The job of etiologic research is to identify components of those etiologies.

## INDUCTION PERIOD

Pie-chart diagrams of sufficient causes and their components such as those in Figure 2–2 are not well suited to provide a model for conceptualizing the *induction period*, which may be defined as the period of time from causal action until disease initiation. There is no way to tell from a pie-chart diagram of a sufficient cause which components affect each other, which components must come before or after others, for which components the temporal order is irrelevant, etc. The crucial

information on temporal ordering must come in a separate description of the interrelations among the components of a sufficient cause.

If, in sufficient cause I, the sequence of action of the specified component causes must be  $A = 0, B = 1$  and we are studying the effect of  $A = 0$ , which (let us assume) acts at a narrowly defined point in time, we do not observe the occurrence of disease immediately after  $A = 0$  occurs. Disease occurs only after the sequence is completed, so there will be a delay while  $B = 1$  occurs (along with components of the set  $U_1$  that are not present or that have not occurred when  $A = 0$  occurs). When  $B = 1$  acts, if it is the last of all the component causes (including those in the set of unspecified conditions and events represented by  $U_1$ ), disease occurs. The interval between the action of  $B = 1$  and the disease occurrence is the induction time for the effect of  $B = 1$  in sufficient cause I.

In the example given earlier of an equilibrium disorder leading to a later fall and hip injury, the induction time between the start of the equilibrium disorder and the later hip injury might be long, if the equilibrium disorder is caused by an old head injury, or short, if the disorder is caused by inebriation. In the latter case, it could even be instantaneous, if we define it as blood alcohol greater than a certain level. This latter possibility illustrates an important general point: Component causes that do not change with time, as opposed to events, all have induction times of zero.

Defining an induction period of interest is tantamount to specifying the characteristics of the component causes of interest. A clear example of a lengthy induction time is the cause-effect relation between exposure of a female fetus to diethylstilbestrol (DES) and the subsequent development of adenocarcinoma of the vagina. The cancer is usually diagnosed between ages 15 and 30 years. Because the causal exposure to DES occurs early in pregnancy, there is an induction time of about 15 to 30 years for the carcinogenic action of DES. During this time, other causes presumably are operating; some evidence suggests that hormonal action during adolescence may be part of the mechanism (Rothman, 1981).

It is incorrect to characterize a disease itself as having a lengthy or brief induction period. The induction time can be conceptualized only in relation to a specific component cause operating in a specific sufficient cause. Thus, we say that the induction time relating DES to clear-cell carcinoma of the vagina is 15 to 30 years, but we should not say that 15 to 30 years is the induction time for clear-cell carcinoma in general. Because each component cause in any causal mechanism can act at a time different from the other component causes, each can have its own induction time. For the component cause that acts last, the induction time equals zero. If another component cause of clear-cell carcinoma of the vagina that acts during adolescence were identified, it would have a much shorter induction time for its carcinogenic action than DES. Thus, induction time characterizes a specific cause-effect pair rather than just the effect.

In carcinogenesis, the terms *initiator* and *promotor* have been used to refer to some of the component causes of cancer that act early and late, respectively, in the causal mechanism. Cancer itself has often been characterized as a disease process with a long induction time. This characterization is a misconception, however, because any late-acting component in the causal process, such as a promotor, will have a short induction time. Indeed, by definition, the induction time will always be zero for at least one component cause, the last to act. The mistaken view that diseases, as opposed to cause-disease relationships, have long or short induction periods can have important implications for research. For instance, the view of adult cancers as "diseases of long latency" may induce some researchers to ignore evidence of etiologic effects occurring relatively late in the processes that culminate in clinically diagnosed cancers. At the other extreme, the routine disregard for exposures occurring in the first decade or two in studies of occupational carcinogenesis, as a major example, may well have inhibited the discovery of occupational causes with very long induction periods.

Disease, once initiated, will not necessarily be apparent. The time interval between irreversible disease occurrence and detection has been termed the *latent period* (Rothman, 1981), although others have used this term interchangeably with induction period. Still others use *latent period* to mean the total time between causal action and disease detection. We use *induction period* to describe the time from causal action to irreversible disease occurrence and *latent period* to mean the time from disease occurrence to disease detection. The latent period can sometimes be reduced by improved methods of disease detection. The induction period, on the other hand, cannot be reduced by early detection of disease, because disease occurrence marks the end of the induction period. Earlier detection of disease, however, may reduce the apparent induction period (the time between causal action and disease detection), because the time when disease is detected, as a practical matter, is

Notes for section 1

he interrelations among  
ent causes must be  $A =$   
ts at a narrowly defined  
r  $A = 0$  occurs. Disease  
 $B = 1$  occurs (along with  
n  $A = 0$  occurs). When  
n the set of unspecified  
ween the action of  $B =$   
n sufficient cause I.  
r fall and hip injury, the  
ip injury might be long,  
e disorder is caused by  
as blood alcohol greater  
oint: Component causes  
s of zero.

ie characteristics of the  
the cause–effect relation  
ubsequent development  
n ages 15 and 30 years.  
induction time of about  
causes presumably are  
nce may be part of the

of induction period. The  
ent cause operating in a  
to clear-cell carcinoma  
s the induction time for  
usal mechanism can act  
wn induction time. For  
her component cause of  
ed, it would have a much  
ion time characterizes a

refer to some of the com-  
mechanism. Cancer itself  
ie. This characterization  
ausal process, such as a  
tion time will always be  
t diseases, as opposed to  
important implications  
ency” may induce some  
ite in the processes that  
disregard for exposures  
sis, as a major example,  
ng induction periods.

val between irreversible  
(Rothman, 1981), although  
ers use *latent period* to  
*action period* to describe  
od to mean the time from  
be reduced by improved  
not be reduced by early  
duction period. Earlier  
the time between causal  
as a practical matter, is

usually used to mark the time of disease occurrence. Thus, diseases such as slow-growing cancers may appear to have long induction periods with respect to many causes because they have long latent periods. The latent period, unlike the induction period, is a characteristic of the disease and the detection effort applied to the person with the disease.

Although it is not possible to reduce the induction period proper by earlier detection of disease, it may be possible to observe intermediate stages of a causal mechanism. The increased interest in biomarkers such as DNA adducts is an example of attempting to focus on causes more proximal to the disease occurrence or on effects more proximal to cause occurrence. Such biomarkers may nonetheless reflect the effects of earlier-acting agents on the person.

Some agents may have a causal action by shortening the induction time of other agents. Suppose that exposure to factor  $X = 1$  leads to epilepsy after an interval of 10 years, on average. It may be that exposure to a drug,  $Z = 1$ , would shorten this interval to 2 years. Is  $Z = 1$  acting as a catalyst, or as a cause, of epilepsy? The answer is both: A catalyst is a cause. Without  $Z = 1$ , the occurrence of epilepsy comes 8 years later than it comes with  $Z = 1$ , so we can say that  $Z = 1$  causes the onset of the early epilepsy. It is not sufficient to argue that the epilepsy would have occurred anyway. First, it would not have occurred at that time, and the time of occurrence is part of our definition of an event. Second, epilepsy will occur later only if the individual survives an additional 8 years, which is not certain. Not only does agent  $Z = 1$  determine when the epilepsy occurs, it can also determine whether it occurs. Thus, we should call any agent that acts as a catalyst of a causal mechanism, speeding up an induction period for other agents, a cause in its own right. Similarly, any agent that postpones the onset of an event, drawing out the induction period for another agent, is a preventive. It should not be too surprising to equate postponement to prevention: We routinely use such an equation when we employ the euphemism that we “prevent” death, which actually can only be postponed. What we prevent is death at a given time, in favor of death at a later time.

### SCOPE OF THE MODEL

The main utility of this model of sufficient causes and their components lies in its ability to provide a general but practical conceptual framework for causal problems. The attempt to make the proportion of disease attributable to various component causes add to 100% is an example of a fallacy that is exposed by the model (although MacMahon and others were able to invoke yellow shanks and phenylketonuria to expose that fallacy long before the sufficient-component cause model was formally described [MacMahon and Pugh, 1967, 1970]). The model makes it clear that, because of interactions, there is no upper limit to the sum of these proportions. As we shall see in Chapter 5, the epidemiologic evaluation of interactions themselves can be clarified, to some extent, with the help of the model.

Although the model appears to deal qualitatively with the action of component causes, it can be extended to account for dose dependence by postulating a set of sufficient causes, each of which contains as a component a different dose of the agent in question. Small doses might require a larger or rarer set of complementary causes to complete a sufficient cause than that required by large doses (Rothman, 1976a), in which case it is particularly important to specify both sides of the causal contrast. In this way, the model can account for the phenomenon of a shorter induction period accompanying larger doses of exposure, because a smaller set of complementary components would be needed to complete the sufficient cause.

Those who believe that chance must play a role in any complex mechanism might object to the intricacy of this seemingly deterministic model. A probabilistic (stochastic) model could be invoked to describe a dose-response relation, for example, without the need for a multitude of different causal mechanisms. The model would simply relate the dose of the exposure to the probability of the effect occurring. For those who believe that virtually all events contain some element of chance, deterministic causal models may seem to misrepresent the indeterminism of the real world. However, the deterministic model presented here can accommodate “chance”; one way might be to view chance, or at least some part of the variability that we call “chance,” as the result of deterministic events that are beyond the current limits of knowledge or observability.

For example, the outcome of a flip of a coin is usually considered a chance event. In classical mechanics, however, the outcome can in theory be determined completely by the application of physical laws and a sufficient description of the starting conditions. To put it in terms more familiar

to epidemiologists, consider the explanation for why an individual gets lung cancer. One hundred years ago, when little was known about the etiology of lung cancer; a scientist might have said that it was a matter of chance. Nowadays, we might say that the risk depends on how much the individual smokes, how much asbestos and radon the individual has been exposed to, and so on. Nonetheless, recognizing this dependence moves the line of ignorance; it does not eliminate it. One can still ask what determines whether an individual who has smoked a specific amount and has a specified amount of exposure to all the other known risk factors will get lung cancer. Some will get lung cancer and some will not, and if all known risk factors are already taken into account, what is left we might still describe as chance. True, we can explain much more of the variability in lung cancer occurrence nowadays than we formerly could by taking into account factors known to cause it, but at the limits of our knowledge, we still ascribe the remaining variability to what we call chance. In this view, chance is seen as a catchall term for our ignorance about causal explanations.

We have so far ignored more subtle considerations of sources of unpredictability in events, such as chaotic behavior (in which even the slightest uncertainty about initial conditions leads to vast uncertainty about outcomes) and quantum-mechanical uncertainty. In each of these situations, a random (stochastic) model component may be essential for any useful modeling effort. Such components can also be introduced in the above conceptual model by treating unmeasured component causes in the model as random events, so that the causal model based on components of sufficient causes can have random elements. An example is treatment assignment in randomized clinical trials (Poole 2001a).

### OTHER MODELS OF CAUSATION

The sufficient-component cause model is only one of several models of causation that may be useful for gaining insight about epidemiologic concepts (Greenland and Brumback, 2002; Greenland, 2004a). It portrays qualitative causal mechanisms within members of a population, so its fundamental unit of analysis is the causal mechanism rather than a person. Many different sets of mechanisms can lead to the same pattern of disease within a population, so the sufficient-component cause model involves specification of details that are beyond the scope of epidemiologic data. Also, it does not incorporate elements reflecting population distributions of factors or causal sequences, which are crucial to understanding confounding and other biases.

Other models of causation, such as potential-outcome (counterfactual) models and graphical models, provide direct representations of epidemiologic concepts such as confounding and other biases, and can be applied at mechanistic, individual, or population levels of analysis. Potential-outcome models (Chapters 4 and 5) specify in detail what would happen to individuals or populations under alternative possible patterns of interventions or exposures, and also bring to the fore problems in operationally defining causes (Greenland, 2002a, 2005a; Hernán, 2005). Graphical models (Chapter 12) display broad qualitative assumptions about causal directions and independencies. Both types of model have close relationships to the structural-equations models that are popular in the social sciences (Pearl, 2000; Greenland and Brumback, 2002), and both can be subsumed under a general theory of longitudinal causality (Robins, 1997).

### PHILOSOPHY OF SCIENTIFIC INFERENCE

Causal inference may be viewed as a special case of the more general process of scientific reasoning. The literature on this topic is too vast for us to review thoroughly, but we will provide a brief overview of certain points relevant to epidemiology, at the risk of some oversimplification.

### INDUCTIVISM

Modern science began to emerge around the 16th and 17th centuries, when the knowledge demands of emerging technologies (such as artillery and transoceanic navigation) stimulated inquiry into the origins of knowledge. An early codification of the scientific method was Francis Bacon's *Novum Organum*, which, in 1620, presented an inductivist view of science. In this philosophy, scientific reasoning is said to depend on making generalizations, or inductions, from observations to general laws of nature; the observations are said to induce the formulation of a natural law in the mind of

cancer. One hundred scientist might have said things on how much the disease is exposed to, and so on. It is not eliminate it. One scientific amount and has a high risk of cancer. Some will be taken into account, some of the variability in important factors known to contribute to what we call causal explanations. Variability in events, such as conditions leads to vast differences in these situations, a great effort. Such components of sufficient optimized clinical trials

tion that may be used (e.g., 2002; Greenland, 2009), so its fundamental sets of mechanisms support cause model data. Also, it does not equences, which are

models and graphical confounding and other analysis. Potential individuals or populations relating to the fore problem. Graphical models and independencies. Those that are popular in are subsumed under

scientific reasoning. Provide a brief overview on.

knowledge demands inquiry into the Francis Bacon's *Novum Organon*, scientific derivations to general law in the mind of

the scientist. Thus, an inductivist would have said that Jenner's observation of lack of smallpox among milkmaids induced in Jenner's mind the theory that cowpox (common among milkmaids) conferred immunity to smallpox. Inductivist philosophy reached a pinnacle of sorts in the canons of John Stuart Mill (1862), which evolved into inferential criteria that are still in use today.

Inductivist philosophy was a great step forward from the medieval scholasticism that preceded it, for at least it demanded that a scientist make careful observations of people and nature rather than appeal to faith, ancient texts, or authorities. Nonetheless, in the 18th century the Scottish philosopher David Hume described a disturbing deficiency in inductivism. An inductive argument carried no logical force; instead, such an argument represented nothing more than an *assumption* that certain events would in the future follow the same pattern as they had in the past. Thus, to argue that cowpox caused immunity to smallpox because no one got smallpox after having cowpox corresponded to an unjustified assumption that the pattern observed to date (no smallpox after cowpox) would continue into the future. Hume pointed out that, even for the most reasonable-sounding of such assumptions, there was no logical necessity behind the inductive argument.

Of central concern to Hume (1739) was the issue of causal inference and failure of induction to provide a foundation for it:

Thus not only our reason fails us in the discovery of the ultimate connexion of causes and effects, but even after experience has inform'd us of their constant conjunction, 'tis impossible for us to satisfy ourselves by our reason, why we shou'd extend that experience beyond those particular instances, which have fallen under our observation. We suppose, but are never able to prove, that there must be a resemblance betwixt those objects, of which we have had experience, and those which lie beyond the reach of our discovery.

In other words, no number of repetitions of a particular sequence of events, such as the appearance of a light after flipping a switch, can prove a causal connection between the action of the switch and the turning on of the light. No matter how many times the light comes on after the switch has been pressed, the possibility of coincidental occurrence cannot be ruled out. Hume pointed out that observers cannot perceive causal connections, but only a series of events. Bertrand Russell (1945) illustrated this point with the example of two accurate clocks that perpetually chime on the hour, with one keeping time slightly ahead of the other. Although one invariably chimes before the other, there is no direct causal connection from one to the other. Thus, assigning a causal interpretation to the pattern of events cannot be a logical extension of our observations alone, because the events might be occurring together only because of a shared earlier cause, or because of some systematic error in the observations.

Causal inference based on mere association of events constitutes a logical fallacy known as *post hoc ergo propter hoc* (Latin for "after this therefore on account of this"). This fallacy is exemplified by the inference that the crowing of a rooster is necessary for the sun to rise because sunrise is always preceded by the crowing.

The *post hoc* fallacy is a special case of a more general logical fallacy known as the *fallacy of affirming the consequent*. This fallacy of confirmation takes the following general form: "We know that if H is true, B must be true; and we know that B is true; therefore H must be true." This fallacy is used routinely by scientists in interpreting data. It is used, for example, when one argues as follows: "If sewer service causes heart disease, then heart disease rates should be highest where sewer service is available; heart disease rates are indeed highest where sewer service is available; therefore, sewer service causes heart disease." Here, H is the hypothesis "sewer service causes heart disease" and B is the observation "heart disease rates are highest where sewer service is available." The argument is logically unsound, as demonstrated by the fact that we can imagine many ways in which the premises could be true but the conclusion false; for example, economic development could lead to both sewer service and elevated heart disease rates, without any effect of sewer service on heart disease. In this case, however, we also know that one of the premises is not true—specifically, the premise, "If H is true, B must be true." This particular form of the fallacy exemplifies the problem of *confounding*, which we will discuss in detail in later chapters.

Bertrand Russell (1945) satirized the fallacy this way:

'If p, then q; now q is true; therefore p is true.' E.g., 'If pigs have wings, then some winged animals are good to eat; now some winged animals are good to eat; therefore pigs have wings.' This form of inference is called 'scientific method.'

### REFUTATIONISM

Russell was not alone in his lament of the illogicality of scientific reasoning as ordinarily practiced. Many philosophers and scientists from Hume's time forward attempted to set out a firm logical basis for scientific reasoning.

In the 1920s, most notable among these was the school of logical positivists, who sought a logic for science that could lead inevitably to correct scientific conclusions, in much the way rigorous logic can lead inevitably to correct conclusions in mathematics. Other philosophers and scientists, however, had started to suspect that scientific hypotheses can never be proven or established as true in any logical sense. For example, a number of philosophers noted that scientific statements can only be found to be consistent with observation, but cannot be proven or disproven in any "airtight" logical or mathematical sense (Duhem, 1906, transl. 1954; Popper 1934, transl. 1959; Quine, 1951). This fact is sometimes called the problem of *nonidentification* or *underdetermination* of theories by observations (Curd and Cover, 1998). In particular, available observations are always consistent with several hypotheses that themselves are mutually inconsistent, which explains why (as Hume noted) scientific theories cannot be logically proven. In particular, consistency between a hypothesis and observations is no proof of the hypothesis, because we can always invent alternative hypotheses that are just as consistent with the observations.

In contrast, a valid observation that is inconsistent with a hypothesis implies that the hypothesis as stated is false and so refutes the hypothesis. If you wring the rooster's neck before it crows and the sun still rises, you have disproved that the rooster's crowing is a necessary cause of sunrise. Or consider a hypothetical research program to learn the boiling point of water (Magee, 1985). A scientist who boils water in an open flask and repeatedly measures the boiling point at 100°C will never, no matter how many confirmatory repetitions are involved, prove that 100°C is always the boiling point. On the other hand, merely one attempt to boil the water in a closed flask or at high altitude will refute the proposition that water always boils at 100°C.

According to Popper, science advances by a process of elimination that he called "conjecture and refutation." Scientists form hypotheses based on intuition, conjecture, and previous experience. Good scientists use deductive logic to infer predictions from the hypothesis and then compare observations with the predictions. Hypotheses whose predictions agree with observations are confirmed (Popper used the term "corroborated") only in the sense that they can continue to be used as explanations of natural phenomena. At any time, however, they may be refuted by further observations and might be replaced by other hypotheses that are more consistent with the observations. This view of scientific inference is sometimes called *refutationism* or *falsificationism*. Refutationists consider induction to be a psychologic crutch: Repeated observations did not in fact induce the formulation of a natural law, but only the belief that such a law has been found. For a refutationist, only the psychologic comfort provided by induction explains why it still has advocates.

One way to rescue the concept of induction from the stigma of pure delusion is to resurrect it as a psychologic phenomenon, as Hume and Popper claimed it was, but one that plays a legitimate role in hypothesis formation. The philosophy of conjecture and refutation places no constraints on the origin of conjectures. Even delusions are permitted as hypotheses, and therefore inductively inspired hypotheses, however psychologic, are valid starting points for scientific evaluation. This concession does not admit a logical role for induction in confirming scientific hypotheses, but it allows the process of induction to play a part, along with imagination, in the scientific cycle of conjecture and refutation.

The philosophy of conjecture and refutation has profound implications for the methodology of science. The popular concept of a scientist doggedly assembling evidence to support a favorite thesis is objectionable from the standpoint of refutationist philosophy because it encourages scientists to consider their own pet theories as their intellectual property, to be confirmed, proven, and, when all the evidence is in, cast in stone and defended as natural law. Such attitudes hinder critical evaluation, interchange, and progress. The approach of conjecture and refutation, in contrast, encourages scientists to consider multiple hypotheses and to seek crucial tests that decide between competing hypotheses by falsifying one of them. Because falsification of one or more theories is the goal, there is incentive to depersonalize the theories. Criticism leveled at a theory need not be seen as criticism of the person who proposed it. It has been suggested that the reason why certain fields of science advance rapidly while others languish is that the rapidly advancing fields are propelled by scientists

ordinarily practiced. et out a firm logical

, who sought a logic ch the way rigorous phers and scientists, or established as true ific statements can ven in any “airtight” 1959; Quine, 1951). mination of theories re always consistent lains why (as Hume etween a hypothesis ternal hypotheses

s that the hypothesis before it crows and try cause of sunrise. er (Magee, 1985). A point at 100°C will 100°C is always the osed flask or at high

e called “conjecture previous experience. then compare obser- ations are confirmed to be used as expla- further observations servations. This view futationists consider duce the formulation futationist, only the

sion is to resurrect it at plays a legitimate es no constraints on herefore inductively ific evaluation. This ic hypotheses, but it e scientific cycle of

the methodology of port a favorite thesis ourages scientists to roven, and, when all under critical evalua- contrast, encourages between competing ries is the goal, there t be seen as criticism tain fields of science opelled by scientists

who are busy constructing and testing competing hypotheses; the other fields, in contrast, “are sick by comparison, because they have forgotten the necessity for alternative hypotheses and disproof” (Platt, 1964).

The refutationist model of science has a number of valuable lessons for research conduct, especially of the need to seek alternative explanations for observations, rather than focus on the chimera of seeking scientific “proof” for some favored theory. Nonetheless, it is vulnerable to criticisms that observations (or some would say their interpretations) are themselves laden with theory (sometimes called the *Duhem-Quine thesis*; Curd and Cover, 1998). Thus, observations can never provide the sort of definitive refutations that are the hallmark of popular accounts of refutationism. For example, there may be uncontrolled and even unimagined biases that have made our refutational observations invalid; to claim refutation is to assume as true the unprovable theory that no such bias exists. In other words, not only are theories underdetermined by observations, so are refutations, which are themselves theory-laden. The net result is that logical certainty about either the truth or falsity of an internally consistent theory is impossible (Quine, 1951).

### **CONSENSUS AND NATURALISM**

Some 20th-century philosophers of science, most notably Thomas Kuhn (1962), emphasized the role of the scientific community in judging the validity of scientific theories. These critics of the conjecture-and-refutation model suggested that the refutation of a theory involves making a choice. Every observation is itself dependent on theories. For example, observing the moons of Jupiter through a telescope seems to us like a direct observation, but only because the theory of optics on which the telescope is based is so well accepted. When confronted with a refuting observation, a scientist faces the choice of rejecting either the validity of the theory being tested or the validity of the refuting observation, which itself must be premised on scientific theories that are not certain (Haack, 2003). Observations that are falsifying instances of theories may at times be treated as “anomalies,” tolerated without falsifying the theory in the hope that the anomalies may eventually be explained. An epidemiologic example is the observation that shallow-inhaling smokers had higher lung cancer rates than deep-inhaling smokers. This anomaly was eventually explained when it was noted that lung tissue higher in the lung is more susceptible to smoking-associated lung tumors, and shallowly inhaled smoke tars tend to be deposited higher in the lung (Wald, 1985).

In other instances, anomalies may lead eventually to the overthrow of current scientific doctrine, just as Newtonian mechanics was displaced (remaining only as a first-order approximation) by relativity theory. Kuhn asserted that in every branch of science the prevailing scientific viewpoint, which he termed “normal science,” occasionally undergoes major shifts that amount to scientific revolutions. These revolutions signal a decision of the scientific community to discard the scientific infrastructure rather than to falsify a new hypothesis that cannot be easily grafted onto it. Kuhn and others have argued that the consensus of the scientific community determines what is considered accepted and what is considered refuted.

Kuhn’s critics characterized this description of science as one of an irrational process, “a matter for mob psychology” (Lakatos, 1970). Those who believe in a rational structure for science consider Kuhn’s vision to be a regrettably real description of much of what passes for scientific activity, but not prescriptive for any good science. Although many modern philosophers reject rigid demarcations and formulations for science such as refutationism, they nonetheless maintain that science is founded on reason, albeit possibly informal common sense (Haack, 2003). Others go beyond Kuhn and maintain that attempts to impose a singular rational structure or methodology on science hobbles the imagination and is a prescription for the same sort of authoritarian repression of ideas that scientists have had to face throughout history (Feyerabend, 1975 and 1993).

The philosophic debate about Kuhn’s description of science hinges on whether Kuhn meant to describe only what has happened historically in science or instead what ought to happen, an issue about which Kuhn (1970) has not been completely clear:

Are Kuhn’s [my] remarks about scientific development . . . to be read as descriptions or prescriptions? The answer, of course, is that they should be read in both ways at once. If I have a theory of how and why science works, it must necessarily have implications for the way in which scientists should behave if their enterprise is to flourish.

The idea that science is a sociologic process, whether considered descriptive or normative, is an interesting thesis, as is the idea that from observing how scientists work we can learn about how scientists ought to work. The latter idea has led to the development of *naturalistic* philosophy of science, or “science studies,” which examines scientific developments for clues about what sort of methods scientists need and develop for successful discovery and invention (Callebaut, 1993; Giere, 1999).

Regardless of philosophical developments, we suspect that most epidemiologists (and most scientists) will continue to function as if the following classical view is correct: The ultimate goal of scientific inference is to capture some objective truths about the material world in which we live, and any theory of inference should ideally be evaluated by how well it leads us to these truths. This ideal is impossible to operationalize, however, for if we ever find any ultimate truths, we will have no way of knowing that for certain. Thus, those holding the view that scientific truth is not arbitrary nevertheless concede that our knowledge of these truths will always be tentative. For refutationists, this tentativeness has an asymmetric quality, but that asymmetry is less marked for others. We may believe that we know a theory is false because it consistently fails the tests we put it through, but our tests could be faulty, given that they involve imperfect reasoning and sense perception. Neither can we know that a theory is true, even if it passes every test we can devise, for it may fail a test that is as yet undevised.

Few, if any, would disagree that a theory of inference should be evaluated at least in part by how well it leads us to detect errors in our hypotheses and observations. There are, however, many other inferential activities besides evaluation of hypotheses, such as prediction or forecasting of events, and subsequent attempts to control events (which of course requires causal information). Statisticians rather than philosophers have more often confronted these problems in practice, so it should not be surprising that the major philosophies concerned with these problems emerged from statistics rather than philosophy.

### BAYESIANISM

There is another philosophy of inference that, like most, holds an objective view of scientific truth and a view of knowledge as tentative or uncertain, but that focuses on evaluation of knowledge rather than truth. Like refutationism, the modern form of this philosophy evolved from the writings of 18th-century thinkers. The focal arguments first appeared in a pivotal essay by the Reverend Thomas Bayes (1764), and hence the philosophy is usually referred to as Bayesianism (Howson and Urbach, 1993), and it was the renowned French mathematician and scientist Pierre Simon de Laplace who first gave it an applied statistical format. Nonetheless, it did not reach a complete expression until after World War I, most notably in the writings of Ramsey (1931) and DeFinetti (1937); and, like refutationism, it did not begin to appear in epidemiology until the 1970s (e.g., Cornfield, 1976).

The central problem addressed by Bayesianism is the following: In classical logic, a deductive argument can provide no information about the truth or falsity of a scientific hypothesis unless you can be 100% certain about the truth of the premises of the argument. Consider the logical argument called *modus tollens*: “If H implies B, and B is false, then H must be false.” This argument is logically valid, but the conclusion follows only on the assumptions that the premises “H implies B” and “B is false” are true statements. If these premises are statements about the physical world, we cannot possibly know them to be correct with 100% certainty, because all observations are subject to error. Furthermore, the claim that “H implies B” will often depend on its own chain of deductions, each with its own premises of which we cannot be certain.

For example, if H is “Television viewing causes homicides” and B is “Homicide rates are highest where televisions are most common,” the first premise used in *modus tollens* to test the hypothesis that television viewing causes homicides will be: “If television viewing causes homicides, homicide rates are highest where televisions are most common.” The validity of this premise is doubtful—after all, even if television does cause homicides, homicide rates may be low where televisions are common because of socioeconomic advantages in those areas.

Continuing to reason in this fashion, we could arrive at a more pessimistic state than even Hume imagined. Not only is induction without logical foundation, *deduction* has limited scientific utility because we cannot ensure the truth of all the premises, even if a logical argument is valid.

The Baye  
and puts  
certainty,  
the rules  
logically  
certainty,  
*prior pro*  
may very  
accumula  
priors to  
Because  
supplying  
subjectiv  
truth. No  
philosoph  
scientific  
logic yiel

A con  
goals of s  
called “ce  
however,  
to test an  
for infere

Another  
and they a  
process is  
it often a  
here is he  
is not “tr  
intuitive  
Gilovich.  
biases an  
to “put th  
they have  
degrees c

In any  
an observ  
challengi  
comings  
certain fi  
precise t  
past stud  
Moreove  
prior pro  
same infi

Probab  
of a the  
hypothes  
for pref  
source p  
are “sel  
disease),  
given a h

Even  
the varyi  
data. Suc

DROPPED

ve or normative, is an  
can learn about how  
realistic philosophy of  
clues about what sort  
ion (Callebaut, 1993;

mologists (and most  
ect: The ultimate goal  
orld in which we live,  
is to these truths. This  
e truths, we will have  
c truth is not arbitrary  
ive. For refutationists,  
ed for others. We may  
we put it through, but  
se perception. Neither  
, for it may fail a test

ted at least in part by  
re are, however, many  
tion or forecasting of  
s causal information).  
blems in practice, so it  
oblems emerged from

view of scientific truth  
luation of knowledge  
ived from the writings  
ssay by the Reverend  
Bayesianism (Howson  
ntist Pierre Simon de  
not reach a complete  
(1931) and DeFinetti  
until the 1970s (e.g.,

ical logic, a deductive  
hypothesis unless you  
r the logical argument  
se." This argument is  
emises "H implies B"  
he physical world, we  
s observations are subject  
n chain of deductions,

icide rates are highest  
to test the hypothesis  
s homicides, homicide  
premise is doubtful—  
where televisions are

nistic state than even  
t has limited scientific  
cal argument is valid.

The Bayesian answer to this problem is partial in that it makes a severe demand on the scientist and puts a severe limitation on the results. It says roughly this: If you can assign a degree of certainty, or personal probability, to the premises of your valid argument, you may use any and all the rules of probability theory to derive a certainty for the conclusion, and this certainty will be a logically valid consequence of your original certainties. An inescapable fact is that your concluding certainty, or *posterior probability*, may depend heavily on what you used as initial certainties, or *prior probabilities*. If those initial certainties are not the same as those of a colleague, that colleague may very well assign a certainty to the conclusion different from the one you derived. With the accumulation of consistent evidence, however, the data can usually force even extremely disparate priors to converge into similar posterior probabilities.

Because the posterior probabilities emanating from a Bayesian inference depend on the person supplying the initial certainties and so may vary across individuals, the inferences are said to be subjective. This subjectivity of Bayesian inference is often mistaken for a subjective treatment of truth. Not only is such a view of Bayesianism incorrect, it is diametrically opposed to Bayesian philosophy. The Bayesian approach represents a constructive attempt to deal with the dilemma that scientific laws and facts should not be treated as known with certainty, whereas classic deductive logic yields conclusions only when some law, fact, or connection is asserted with 100% certainty.

A common criticism of Bayesian philosophy is that it diverts attention away from the classic goals of science, such as the discovery of how the world works, toward psychologic states of mind called "certainties," "subjective probabilities," or "degrees of belief" (Popper, 1959). This criticism, however, fails to recognize the importance of a scientist's state of mind in determining what theories to test and what tests to apply, the consequent influence of those states on the store of data available for inference, and the influence of the data on the states of mind.

Another reply to this criticism is that scientists already use data to influence their degrees of belief, and they are not shy about expressing those degrees of certainty. The problem is that the conventional process is informal, intuitive, and ineffable, and therefore not subject to critical scrutiny; at its worst, it often amounts to nothing more than the experts announcing that they have seen the evidence and here is how certain they are. How they reached this certainty is left unclear, or, put another way, is not "transparent." The problem is that no one, even an expert, is very good at informally and intuitively formulating certainties that predict facts and future events well (Kahneman et al., 1982; Gilovich, 1993; Piattelli-Palmarini, 1994; Gilovich et al., 2002). One reason for this problem is that biases and prior prejudices can easily creep into expert judgments. Bayesian methods force experts to "put their cards on the table" and specify explicitly the strength of their prior beliefs and why they have such beliefs, defend those specifications against arguments and evidence, and update their degrees of certainty with new evidence in ways that do not violate probability logic.

In any research context, there will be an unlimited number of hypotheses that could explain an observed phenomenon. Some argue that progress is best aided by severely testing (empirically challenging) those explanations that seem most probable in light of past research, so that shortcomings of currently "received" theories can be most rapidly discovered. Indeed, much research in certain fields takes this form, as when theoretical predictions of particle mass are put to ever more precise tests in physics experiments. This process does not involve mere improved repetition of past studies. Rather, it involves tests of previously untested but important predictions of the theory. Moreover, there is an imperative to make the basis for prior beliefs criticizable and defensible. That prior probabilities can differ among persons does not mean that all such beliefs are based on the same information, nor that all are equally tenable.

Probabilities of auxiliary hypotheses are also important in study design and interpretation. Failure of a theory to pass a test can lead to rejection of the theory more rapidly when the auxiliary hypotheses on which the test depends possess high probability. This observation provides a rationale for preferring "nested" case-control studies (in which controls are selected from a roster of the source population for the cases) to "hospital-based" case-control studies (in which the controls are "selected" by the occurrence or diagnosis of one or more diseases other than the case-defining disease), because the former have fewer mechanisms for biased subject selection and hence are given a higher probability of unbiased subject selection.

Even if one disputes the above arguments, most epidemiologists desire some way of expressing the varying degrees of certainty about possible values of an effect measure in light of available data. Such expressions must inevitably be derived in the face of considerable uncertainty about

— P A I S      O T E R S      + V I A E R A —

methodologic details and various events that led to the available data and can be extremely sensitive to the reasoning used in its derivation. For example, as we shall discuss at greater length in Chapter 19, conventional confidence intervals quantify only random error under often questionable assumptions and so should not be interpreted as measures of total uncertainty, particularly for nonexperimental studies. As noted earlier, most people, including scientists, reason poorly in the face of uncertainty. At the very least, subjective Bayesian philosophy provides a methodology for sound reasoning under uncertainty and, in particular, provides many warnings against being overly certain about one's conclusions (Greenland 1998a, 1988b, 2006a; see also Chapters 18 and 19).

Such warnings are echoed in refutationist philosophy. As Peter Medawar (1979) put it, "I cannot give any scientist of any age better advice than this: the intensity of the conviction that a hypothesis is true has no bearing on whether it is true or not." We would add two points. First, the intensity of conviction that a hypothesis is false has no bearing on whether it is false or not. Second, Bayesian methods do not mistake beliefs for evidence. They use evidence to modify beliefs, which scientists routinely do in any event, but often in implicit, intuitive, and incoherent ways.

### **IMPOSSIBILITY OF SCIENTIFIC PROOF**

Vigorous debate is a characteristic of modern scientific philosophy, no less in epidemiology than in other areas (Rothman, 1988). Can divergent philosophies of science be reconciled? Haack (2003) suggested that the scientific enterprise is akin to solving a vast, collective crossword puzzle. In areas in which the evidence is tightly interlocking, there is more reason to place confidence in the answers, but in areas with scant information, the theories may be little better than informed guesses. Of the scientific method, Haack (2003) said that "there is less to the 'scientific method' than meets the eye. Is scientific inquiry categorically different from other kinds? No. Scientific inquiry is continuous with everyday empirical inquiry—only more so."

Perhaps the most important common thread that emerges from the debated philosophies is that proof is impossible in empirical science. This simple fact is especially important to observational epidemiologists, who often face the criticism that proof is impossible in epidemiology, with the implication that it is possible in other scientific disciplines. Such criticism may stem from a view that experiments are the definitive source of scientific knowledge. That view is mistaken on at least two counts. First, the nonexperimental nature of a science does not preclude impressive scientific discoveries; the myriad examples include plate tectonics, the evolution of species, planets orbiting other stars, and the effects of cigarette smoking on human health. Even when they are possible, experiments (including randomized trials) do not provide anything approaching proof and in fact may be controversial, contradictory, or nonreproducible. If randomized clinical trials provided proof, we would never need to do more than one of them on a given hypothesis. Neither physical nor experimental science is immune to such problems, as demonstrated by episodes such as the experimental "discovery" (later refuted) of cold fusion (Taubes, 1993).

Some experimental scientists hold that epidemiologic relations are only suggestive and believe that detailed laboratory study of mechanisms within single individuals can reveal cause–effect relations with certainty. This view overlooks the fact that *all* relations are suggestive in exactly the manner discussed by Hume. Even the most careful and detailed mechanistic dissection of individual events cannot provide more than associations, albeit at a finer level. Laboratory studies often involve a degree of observer control that cannot be approached in epidemiology; it is only this control, not the level of observation, that can strengthen the inferences from laboratory studies. And again, such control is no guarantee against error. In addition, neither scientists nor decision makers are often highly persuaded when only mechanistic evidence from the laboratory is available.

All of the fruits of scientific work, in epidemiology or other disciplines, are at best only tentative formulations of a description of nature, even when the work itself is carried out without mistakes. The tentativeness of our knowledge does not prevent practical applications, but it should keep us skeptical and critical, not only of everyone else's work, but of our own as well. Sometimes etiologic hypotheses enjoy an extremely high, universally or almost universally shared, degree of certainty. The hypothesis that cigarette smoking causes lung cancer is one of the best-known examples. These hypotheses rise above "tentative" acceptance and are the closest we can come to "proof." But even

these hy  
of a mat

### **CAUSALITY**

Etiologic  
selves at  
such as  
sequence  
on testin  
not have  
hypothe  
not solel

### **TESTS**

If the ca  
crucial t  
was first  
it was a  
from the  
that proc  
use. The  
frequenc  
agent, m  
absorptio  
more fre  
culture-i  
lower ra  
each tan  
dose of i  
associat  
was, how  
their per  
about th  
and Rotl

Anot  
observat  
of endor  
associati  
that indu  
endometr  
estrogen  
that wit  
at all. M  
detectio  
have the  
medical  
estrogen  
the detec  
cancer c  
many ot  
huge res  
the muc  
observat

in be extremely sensi-  
ss at greater length in  
under often question-  
certainty, particularly  
tists, reason poorly in  
vides a methodology  
arnings against being  
see also Chapters 18

1979) put it, "I cannot  
tion that a hypothesis  
First, the intensity of  
ot. Second, Bayesian  
iefs, which scientists  
s.

epidemiology than in  
nciled? Haack (2003)  
word puzzle. In areas  
idence in the answers,  
rmed guesses. Of the  
d' than meets the eye.  
inquiry is continuous

d philosophies is that  
rtant to observational  
epidemiology, with the  
ay stem from a view  
s mistaken on at least  
impressive scientific  
cies, planets orbiting  
en they are possible,  
ing proof and in fact  
nical trials provided  
sis. Neither physical  
episodes such as the

iggestive and believe  
l reveal cause-effect  
gestive in exactly the  
section of individual  
studies often involve  
only this control, not  
lies. And again, such  
ion makers are often  
able.  
at best only tentative  
ut without mistakes.  
ut it should keep us  
Sometimes etiologic  
, degree of certainty.  
own examples. These  
to "proof." But even

these hypotheses are not "proved" with the degree of absolute certainty that accompanies the proof of a mathematical theorem.

## CAUSAL INFERENCE IN EPIDEMIOLOGY

Etiologic knowledge about epidemiologic hypotheses is often scant, making the hypotheses themselves at times little more than vague statements of causal association between exposure and disease, such as "smoking causes cardiovascular disease." These vague hypotheses have only vague consequences that can be difficult to test. To cope with this vagueness, epidemiologists usually focus on testing the negation of the causal hypothesis, that is, the null hypothesis that the exposure does *not* have a causal relation to disease. Then, any observed association can potentially refute the hypothesis, subject to the assumption (auxiliary hypothesis) that biases and chance fluctuations are not solely responsible for the observation.

## TESTS OF COMPETING EPIDEMIOLOGIC THEORIES

If the causal mechanism is stated specifically enough, epidemiologic observations can provide crucial tests of competing, non-null causal hypotheses. For example, when toxic-shock syndrome was first studied, there were two competing hypotheses about the causal agent. Under one hypothesis, it was a chemical in the tampon, so that women using tampons were exposed to the agent directly from the tampon. Under the other hypothesis, the tampon acted as a culture medium for staphylococci that produced a toxin. Both hypotheses explained the relation of toxic-shock occurrence to tampon use. The two hypotheses, however, led to opposite predictions about the relation between the frequency of changing tampons and the rate of toxic shock. Under the hypothesis of a chemical agent, more frequent changing of the tampon would lead to more exposure to the agent and possible absorption of a greater overall dose. This hypothesis predicted that women who changed tampons more frequently would have a higher rate than women who changed tampons infrequently. The culture-medium hypothesis predicts that women who change tampons frequently would have a lower rate than those who change tampons less frequently, because a short duration of use for each tampon would prevent the staphylococci from multiplying enough to produce a damaging dose of toxin. Thus, epidemiologic research, by showing that infrequent changing of tampons was associated with a higher rate of toxic shock, refuted the chemical theory in the form presented. There was, however, a third hypothesis that a chemical in some tampons (e.g., oxygen content) improved their performance as culture media. This chemical-promotor hypothesis made the same prediction about the association with frequency of changing tampons as the microbial toxin hypothesis (Lanes and Rothman, 1990).

Another example of a theory that can be easily tested by epidemiologic data relates to the observation that women who took replacement estrogen therapy had a considerably elevated rate of endometrial cancer. Horwitz and Feinstein (1978) conjectured a competing theory to explain the association: They proposed that women taking estrogen experienced symptoms such as bleeding that induced them to consult a physician. The resulting diagnostic workup led to the detection of endometrial cancer at an earlier stage in these women, as compared with women who were not taking estrogens. Horwitz and Feinstein argued that the association arose from this detection bias, claiming that without the bleeding-induced workup, many of these cancers would not have been detected at all. Many epidemiologic observations were used to evaluate these competing hypotheses. The detection-bias theory predicted that women who had used estrogens for only a short time would have the greatest elevation in their rate, as the symptoms related to estrogen use that led to the medical consultation tended to appear soon after use began. Because the association of recent estrogen use and endometrial cancer was the same in both long- and short-term estrogen users, the detection-bias theory was refuted as an explanation for all but a small fraction of endometrial cancer cases occurring after estrogen use. Refutation of the detection-bias theory also depended on many other observations. Especially important was the theory's implication that there must be a huge reservoir of undetected endometrial cancer in the typical population of women to account for the much greater rate observed in estrogen users, an implication that was not borne out by further observations (Hutchison and Rothman, 1978).

MOROCCO Example —

The endometrial cancer example illustrates a critical point in understanding the process of causal inference in epidemiologic studies: Many of the hypotheses being evaluated in the interpretation of epidemiologic studies are auxiliary hypotheses in the sense that they are independent of the presence, absence, or direction of any causal connection between the study exposure and the disease. For example, explanations of how specific types of bias could have distorted an association between exposure and disease are the usual alternatives to the primary study hypothesis. Much of the interpretation of epidemiologic studies amounts to the testing of such auxiliary explanations for observed associations.

### **CAUSAL CRITERIA**

In practice, how do epidemiologists separate causal from noncausal explanations? Despite philosophic criticisms of inductive inference, inductively oriented considerations are often used as criteria for making such inferences (Weed and Gorelic, 1996). If a set of necessary and sufficient causal criteria could be used to distinguish causal from noncausal relations in epidemiologic studies, the job of the scientist would be eased considerably. With such criteria, all the concerns about the logic or lack thereof in causal inference could be subsumed: It would only be necessary to consult the checklist of criteria to see if a relation were causal. We know from the philosophy reviewed earlier that a set of sufficient criteria does not exist. Nevertheless, lists of causal criteria have become popular, possibly because they seem to provide a road map through complicated territory, and perhaps because they suggest hypotheses to be evaluated in a given problem.

A commonly used set of criteria was based on a list of considerations or “viewpoints” proposed by Sir Austin Bradford Hill (1965). Hill’s list was an expansion of a list offered previously in the landmark U.S. Surgeon General’s report *Smoking and Health* (1964), which in turn was anticipated by the inductive canons of John Stuart Mill (1862) and the rules given by Hume (1739). Subsequently, others, especially Susser, have further developed causal considerations (Kaufman and Poole, 2000).

Hill suggested that the following considerations in attempting to distinguish causal from non-causal associations that were already “perfectly clear-cut and beyond what we would care to attribute to the play of chance”: (1) strength, (2) consistency, (3) specificity, (4) temporality, (5) biologic gradient, (6) plausibility, (7) coherence, (8) experimental evidence, and (9) analogy. Hill emphasized that causal inferences cannot be based on a set of rules, condemned emphasis on statistical significance testing, and recognized the importance of many other factors in decision making (Phillips and Goodman, 2004). Nonetheless, the misguided but popular view that his considerations should be used as criteria for causal inference makes it necessary to examine them in detail.

#### **Strength**

Hill argued that strong associations are particularly compelling because, for weaker associations, it is “easier” to imagine what today we would call an unmeasured confounder that might be responsible for the association. Several years earlier, Cornfield et al. (1959) drew similar conclusions. They concentrated on a single hypothetical confounder that, by itself, would explain entirely an observed association. They expressed a strong preference for ratio measures of strength, as opposed to difference measures, and focused on how the observed estimate of a risk ratio provides a minimum for the association that a completely explanatory confounder must have with the exposure (rather than a minimum for the confounder–disease association). Of special importance, Cornfield et al. acknowledged that having only a weak association does not rule out a causal connection (Rothman and Poole, 1988). Today, some associations, such as those between smoking and cardiovascular disease or between environmental tobacco smoke and lung cancer, are accepted by most as causal even though the associations are considered weak.

Counterexamples of strong but noncausal associations are also not hard to find; any study with strong confounding illustrates the phenomenon. For example, consider the strong relation between Down syndrome and birth rank, which is confounded by the relation between Down syndrome and maternal age. Of course, once the confounding factor is identified, the association is diminished by controlling for the factor.

These examples remind us that a strong association is neither necessary nor sufficient for causality, and that weakness is neither necessary nor sufficient for absence of causality. A strong association

bears on  
or other

#### **Consist**

To most  
populati  
associat  
More pi  
causes &  
be met.  
do not a  
but only  
ent only  
seldom.  
ferent ri  
only to  
studies.

One  
mention  
some re  
fallacio  
from a  
differen  
studies.  
icant in  
not othe  
signific  
ation of  
worsen  
(Greenl

#### **Specifi**

The crit  
effects.  
The for  
of the a  
ratio co  
were e:  
cancer.  
several  
with sn  
that the  
howeve  
justific.  
associa

Wei  
noncau  
relatio  
one ha  
the exa  
to 70%  
the rea  
If the e  
suppor  
those r

ng the process of causal ed in the interpretation are independent of the / exposure and the distorted an association ly hypothesis. Much of xiliary explanations for

iations? Despite philo- re often used as criteria y and sufficient causal iemologic studies, the concerns about the logic ecessary to consult the sophy reviewed earlier eria have become pop- l territory, and perhaps

ewpoints" proposed by previously in the land- urn was anticipated by (1739). Subsequently, (Kaufman and Poole,

uish causal from non- would care to attribute rality, (5) biologic gra- logy. Hill emphasized is on statistical signif- ision making (Phillips considerations should in detail.

eaker associations, it is t might be responsible lar conclusions. They n entirely an observed length, as opposed to o provides a minimum h the exposure (rather tance, Cornfield et al. connection (Rothman ng and cardiovascular ted by most as causal

o find; any study with rong relation between 1 Down syndrome and ation is diminished by

r sufficient for causal- y. A strong association

bears only on hypotheses that the association is entirely or partially due to unmeasured confounders or other source of modest bias.

### **Consistency**

To most observers, consistency refers to the repeated observation of an association in different populations under different circumstances. Lack of consistency, however, does not rule out a causal association, because some effects are produced by their causes only under unusual circumstances. More precisely, the effect of a causal agent cannot occur unless the complementary component causes act or have already acted to complete a sufficient cause. These conditions will not always be met. Thus, transfusions can cause infection with the human immunodeficiency virus, but they do not always do so: The virus must also be present. Tampon use can cause toxic-shock syndrome, but only rarely, when certain other, perhaps unknown, conditions are met. Consistency is apparent only after all the relevant details of a causal mechanism are understood, which is to say very seldom. Furthermore, even studies of exactly the same phenomena can be expected to yield different results simply because they differ in their methods and random errors. Consistency serves only to rule out hypotheses that the association is attributable to some factor that varies across studies.

One mistake in implementing the consistency criterion is so common that it deserves special mention. It is sometimes claimed that a literature or set of results is inconsistent simply because some results are "statistically significant" and some are not. This sort of evaluation is completely fallacious even if one accepts the use of significance testing methods. The results (effect estimates) from a set of studies could all be identical even if many were significant and many were not, the difference in significance arising solely because of differences in the standard errors or sizes of the studies. Conversely, the results could be significantly in conflict even if all were all were nonsignificant individually, simply because in aggregate an effect could be apparent in some subgroups but not others (see Chapter 33). The fallacy of judging consistency by comparing *P*-values or statistical significance is not eliminated by "standardizing" estimates (i.e., dividing them by the standard deviation of the outcome, multiplying them by the standard deviation of the exposure, or both); in fact it is worsened, as such standardization can create differences where none exists, or mask true differences (Greenland et al., 1986, 1991; see Chapters 21 and 33).

### **Specificity**

The criterion of specificity has two variants. One is that a cause leads to a single effect, not multiple effects. The other is that an effect has one cause, not multiple causes. Hill mentioned both of them. The former criterion, specificity of effects, was used as an argument in favor of a causal interpretation of the association between smoking and lung cancer and, in an act of circular reasoning, in favor of ratio comparisons and not differences as the appropriate measures of strength. When ratio measures were examined, the association of smoking to diseases looked "quantitatively specific" to lung cancer. When difference measures were examined, the association appeared to be nonspecific, with several diseases (other cancers, coronary heart disease, etc.) being at least as strongly associated with smoking as lung cancer was. Today we know that smoking affects the risk of many diseases and that the difference comparisons were accurately portraying this lack of specificity. Unfortunately, however, the historical episode of the debate over smoking and health is often cited today as justification for the specificity criterion and for using ratio comparisons to measure strength of association. The proper lessons to learn from that episode should be just the opposite.

Weiss (2002) argued that specificity can be used to distinguish some causal hypotheses from noncausal hypotheses, when the causal hypothesis predicts a relation with one outcome but no relation with another outcome. His argument is persuasive when, in addition to the causal hypothesis, one has an alternative noncausal hypothesis that predicts a nonspecific association. Weiss offered the example of screening sigmoidoscopy, which was associated in case-control studies with a 50% to 70% reduction in mortality from distal tumors of the rectum and tumors of the distal colon, within the reach of the sigmoidoscope, but no reduction in mortality from tumors elsewhere in the colon. If the effect of screening sigmoidoscopy were not specific to the distal colon tumors, it would lend support not to all noncausal theories to explain the association, as Weiss suggested, but only to those noncausal theories that would have predicted a nonspecific association. Thus, specificity can

come into play when it can be logically deduced from the causal hypothesis in question and when nonspecificity can be logically deduced from one or more noncausal hypotheses.

### **Temporality**

Temporality refers to the necessity that the cause precede the effect in time. This criterion is inarguable, insofar as any claimed observation of causation must involve the putative cause C preceding the putative effect D. It does *not*, however, follow that a reverse time order is evidence against the hypothesis that C can cause D. Rather, observations in which C followed D merely show that C could not have caused D in these instances; they provide no evidence for or against the hypothesis that C can cause D in those instances in which it precedes D. Only if it is found that C cannot precede D can we dispense with the causal hypothesis that C *could* cause D.

### **Biologic Gradient**

Biologic gradient refers to the presence of a dose-response or exposure-response curve with an expected shape. Although Hill referred to a "linear" gradient, without specifying the scale, a linear gradient on one scale, such as the risk, can be distinctly nonlinear on another scale, such as the log risk, the odds, or the log odds. We might relax the expectation from linear to strictly monotonic (steadily increasing or decreasing) or even further merely to monotonic (a gradient that never changes direction). For example, more smoking means more carcinogen exposure and more tissue damage, hence more opportunity for carcinogenesis. Some causal associations, however, show a rapid increase in response (an approximate threshold effect) rather than a strictly monotonic trend. An example is the association between DES and adenocarcinoma of the vagina. A possible explanation is that the doses of DES that were administered were all sufficiently great to produce the maximum effect from DES. Under this hypothesis, for all those exposed to DES, the development of disease would depend entirely on other component causes.

The somewhat controversial topic of alcohol consumption and mortality is another example. Death rates are higher among nondrinkers than among moderate drinkers, but they ascend to the highest levels for heavy drinkers. There is considerable debate about which parts of the J-shaped dose-response curve are causally related to alcohol consumption and which parts are noncausal artifacts stemming from confounding or other biases. Some studies appear to find only an increasing relation between alcohol consumption and mortality, possibly because the categories of alcohol consumption are too broad to distinguish different rates among moderate drinkers and nondrinkers, or possibly because they have less confounding at the lower end of the consumption scale.

Associations that do show a monotonic trend in disease frequency with increasing levels of exposure are not necessarily causal. Confounding can result in a monotonic relation between a noncausal risk factor and disease if the confounding factor itself demonstrates a biologic gradient in its relation with disease. The relation between birth rank and Down syndrome mentioned earlier shows a strong biologic gradient that merely reflects the progressive relation between maternal age and occurrence of Down syndrome.

These issues imply that the existence of a monotonic association is neither necessary nor sufficient for a causal relation. A nonmonotonic relation only refutes those causal hypotheses specific enough to predict a monotonic dose-response curve.

### **Plausibility**

Plausibility refers to the scientific plausibility of an association. More than any other criterion, this one shows how narrowly systems of causal criteria are focused on epidemiology. The starting point is an epidemiologic association. In asking whether it is causal or not, one of the considerations we take into account is its plausibility. From a less parochial perspective, the entire enterprise of causal inference would be viewed as the act of determining how plausible a causal *hypothesis* is. One of the considerations we would take into account would be epidemiologic associations, if they are available. Often they are not, but causal inference must be done nevertheless, with inputs from toxicology, pharmacology, basic biology, and other sciences.

Just as epidemiology is not essential for causal inference, plausibility can change with the times. Sartwell (1960) emphasized this point, citing remarks of Cheever in 1861, who had been commenting on the etiology of typhus before its mode of transmission (via body lice) was known:

s in question and when  
ieses.

time. This criterion is  
the putative cause C  
time order is evidence  
C followed D merely  
lence for or against the  
ly if it is found that C  
cause D.

esponse curve with an  
ing the scale, a linear  
r scale, such as the log  
t to strictly monotonic  
(a gradient that never  
posure and more tissue  
ations, however, show  
a strictly monotonic  
the vagina. A possible  
ly great to produce the  
DES, the development

ty is another example,  
but they ascend to the  
1 parts of the J-shaped  
ch parts are noncausal  
ar to find only an in-  
cause the categories of  
noderate drinkers and  
nd of the consumption

h increasing levels of  
nic relation between a  
tes a biologic gradient  
ome mentioned earlier  
between maternal age

necessary nor sufficient  
theses specific enough

ny other criterion, this  
ogy. The starting point  
of the considerations  
he entire enterprise of  
a causal *hypothesis* is.  
ic associations, if they  
less, with inputs from

can change with the  
1 1861, who had been  
body lice) was known:

It could be no more ridiculous for the stranger who passed the night in the steerage of an emigrant ship to ascribe the typhus, which he there contracted, to the vermin with which bodies of the sick might be infested. An adequate cause, one reasonable in itself, must correct the coincidences of simple experience.

What was to Cheever an implausible explanation turned out to be the correct explanation, because it was indeed the vermin that caused the typhus infection. Such is the problem with plausibility: It is too often based not on logic or data, but only on prior beliefs. This is not to say that biologic knowledge should be discounted when a new hypothesis is being evaluated, but only to point out the difficulty in applying that knowledge.

The Bayesian approach to inference attempts to deal with this problem by requiring that one quantify, on a probability (0 to 1) scale, the certainty that one has in prior beliefs, as well as in new hypotheses. This quantification displays the dogmatism or open-mindedness of the analyst in a public fashion, with certainty values near 1 or 0 betraying a strong commitment of the analyst for or against a hypothesis. It can also provide a means of testing those quantified beliefs against new evidence (Howson and Urbach, 1993). Nevertheless, no approach can transform plausibility into an objective causal criterion.

### **Coherence**

Taken from the U.S. Surgeon General's *Smoking and Health* (1964), the term *coherence* implies that a cause-and-effect interpretation for an association does not conflict with what is known of the natural history and biology of the disease. The examples Hill gave for coherence, such as the histopathologic effect of smoking on bronchial epithelium (in reference to the association between smoking and lung cancer) or the difference in lung cancer incidence by sex, could reasonably be considered examples of plausibility, as well as coherence; the distinction appears to be a fine one. Hill emphasized that the absence of coherent information, as distinguished, apparently, from the presence of conflicting information, should not be taken as evidence against an association being considered causal. On the other hand, the presence of conflicting information may indeed refute a hypothesis, but one must always remember that the conflicting information may be mistaken or misinterpreted. An example mentioned earlier is the "inhalation anomaly" in smoking and lung cancer, the fact that the excess of lung cancers seen among smokers seemed to be concentrated at sites in the upper airways of the lung. Several observers interpreted this anomaly as evidence that cigarettes were not responsible for the excess. Other observations, however, suggested that cigarette-borne carcinogens were deposited preferentially where the excess was observed, and so the anomaly was in fact consistent with a causal role for cigarettes (Wald, 1985).

### **Experimental Evidence**

To different observers, experimental evidence can refer to clinical trials, to laboratory experiments with rodents or other nonhuman organisms, or to both. Evidence from human experiments, however, is seldom available for epidemiologic research questions, and animal evidence relates to different species and usually to levels of exposure very different from those that humans experience. Uncertainty in extrapolations from animals to humans often dominates the uncertainty of quantitative risk assessments (Freedman and Zeisel, 1988; Crouch et al., 1997).

To Hill, however, experimental evidence meant something else: the "experimental, or semi-experimental evidence" obtained from reducing or eliminating a putatively harmful exposure and seeing if the frequency of disease subsequently declines. He called this the strongest possible evidence of causality that can be obtained. It can be faulty, however, as the "semi-experimental" approach is nothing more than a "before-and-after" time trend analysis, which can be confounded or otherwise biased by a host of concomitant secular changes. Moreover, even if the removal of exposure does causally reduce the frequency of disease, it might not be for the etiologic reason hypothesized. The draining of a swamp near a city, for instance, would predictably and causally reduce the rate of yellow fever or malaria in that city the following summer. But it would be a mistake to call this observation the strongest possible evidence of a causal role of miasmas (Poole, 1999).

### Analogy

Whatever insight might be derived from analogy is handicapped by the inventive imagination of scientists who can find analogies everywhere. At best, analogy provides a source of more elaborate hypotheses about the associations under study; absence of such analogies reflects only lack of imagination or experience, not falsity of the hypothesis.

We might find naive Hill's examples in which reasoning by analogy from the thalidomide and rubella tragedies made it more likely to him that other medicines and infections might cause other birth defects. But such reasoning is common; we suspect most people find it more credible that smoking might cause, say, stomach cancer, because of its associations, some widely accepted as causal, with cancers in other internal and gastrointestinal organs. Here we see how the analogy criterion can be at odds with either of the two specificity criteria. The more apt the analogy, the less specific are the effects of a cause or the less specific the causes of an effect.

### Summary

As is evident, the standards of epidemiologic evidence offered by Hill are saddled with reservations and exceptions. Hill himself was ambivalent about their utility. He did not use the word *criteria* in the speech. He called them "viewpoints" or "perspectives." On the one hand, he asked, "In what circumstances can we pass from this observed *association* to a verdict of *causation?*" (emphasis in original). Yet, despite speaking of verdicts on causation, he disagreed that any "hard-and-fast rules of evidence" existed by which to judge causation: "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).

Actually, as noted above, the fourth viewpoint, temporality, is a *sine qua non* for causal explanations of observed associations. Nonetheless, it does not bear on the hypothesis that an exposure is capable of causing a disease in situations as yet unobserved (whether in the past or the future). For suppose every exposed case of disease ever reported had received the exposure after developing the disease. This reversed temporal relation would imply that exposure had not caused disease among these reported cases, and thus would refute the hypothesis that it had. Nonetheless, it would not refute the hypothesis that the exposure is *capable* of causing the disease, or that it had caused the disease in unobserved cases. It would mean only that we have no worthwhile epidemiologic evidence relevant to that hypothesis, for we had not yet seen what became of those exposed before disease occurred relative to those unexposed. Furthermore, what appears to be a causal sequence could represent reverse causation if preclinical symptoms of the disease lead to exposure, and then overt disease follows, as when patients in pain take analgesics, which may be the result of disease that is later diagnosed, rather than a cause.

Other than temporality, there is no necessary or sufficient criterion for determining whether an observed association is causal. Only when a causal hypothesis is elaborated to the extent that one can predict from it a particular form of consistency, specificity, biologic gradient, and so forth, can "causal criteria" come into play in evaluating causal hypotheses, and even then they do not come into play in evaluating the general hypothesis per se, but only some specific causal hypotheses, leaving others untested.

This conclusion accords with the views of Hume and many others that causal inferences cannot attain the certainty of logical deductions. Although some scientists continue to develop causal considerations as aids to inference (Susser, 1991), others argue that it is detrimental to cloud the inferential process by considering checklist criteria (Lanes and Poole, 1984). An intermediate, refutationist approach seeks to transform proposed criteria into deductive tests of causal hypotheses (Maclure, 1985; Weed, 1986). Such an approach helps avoid the temptation to use causal criteria simply to buttress pet theories at hand, and instead allows epidemiologists to focus on evaluating competing causal theories using crucial observations. Although this refutationist approach to causal inference may seem at odds with the common implementation of Hill's viewpoints, it actually seeks to answer the fundamental question posed by Hill, and the ultimate purpose of the viewpoints he promulgated:

What [the nine viewpoints] can do, with greater or less strength, is to help us to make up our minds on the fundamental question—is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect? (Hill, 1965)

ventive imagination of  
urce of more elaborate  
s reflects only lack of

m the thalidomide and  
ions might cause other  
1 it more credible that  
ne widely accepted as  
see how the analogy  
pt the analogy, the less

Idled with reservations  
se the word *criteria* in  
id, he asked, "In what  
*causation?*" (emphasis  
that any "hard-and-fast  
viewpoints can bring  
ne can be required as a

*i non* for causal expla-  
hesis that an exposure  
the past or the future).  
posure after developing  
ad not caused disease  
Nonetheless, it would  
, or that it had caused  
hwhile epidemiologic  
f those exposed before  
be a causal sequence  
l to exposure, and then  
e the result of disease

determining whether an  
l to the extent that one  
lient, and so forth, can  
hen they do not come  
fic causal hypotheses,

usual inferences cannot  
to develop causal con-  
al to cloud the inferen-  
rmediate, refutationist  
hypotheses (MacLure,  
usual criteria simply to  
evaluating competing  
ich to causal inference  
ctually seeks to answer  
points he promulgated:

to make up our minds  
facts before us, is there

The crucial phrase "equally or more likely than cause and effect" suggests to us a subjective assessment of the certainty, or probability of the causal hypothesis at issue relative to another hypothesis. Although Hill wrote at a time when expressing uncertainty as a probability was unpopular in statistics, it appears from his statement that, for him, causal inference is a subjective matter of degree of personal belief, certainty, or conviction. In any event, this view is precisely that of subjective Bayesian statistics (Chapter 18).

It is unsurprising that case studies (e.g., Weed and Gorelick, 1996) and surveys of epidemiologists (Holman et al., 2001) show, contrary to the rhetoric that often attends invocations of causal criteria, that epidemiologists have *not* agreed on a set of causal criteria or on how to apply them. In one study in which epidemiologists were asked to employ causal criteria to fictional summaries of epidemiologic literatures, the agreement was only slightly greater than would have been expected by chance (Holman et al., 2001). The typical use of causal criteria is to make a case for a position for or against causality that has been arrived at by other, unstated means. Authors pick and choose among the criteria they deploy, and define and weight them in *ad hoc* ways that depend only on the exigencies of the discussion at hand. In this sense, causal criteria appear to function less like standards or principles and more like values (Poole, 2001b), which vary across individual scientists and even vary within the work of a single scientist, depending on the context and time. Thus universal and objective causal criteria, if they exist, have yet to be identified.