

- Subjects No. 32, Ministry of Health. London: His Majesty's Stationery Office; 1926.
18. Winkelstein W Jr. Janet Lane-Clayton, a forgotten epidemiologic pioneer. *Epidemiology*. 2006;17:705.
  19. Paneth N, Susser E, Susser M. Origins and early development of the case-control study. Part 2: The case-control study from Lane-Clayton to 1950. *Social Prev Med*. 2002;47:359-365.
  20. Daniel TM. *Wade Hampton Frost, Pioneer Epidemiologist 1880-1938*. Rochester, NY: University of Rochester Press; 2004.

## What Is Causation?

The acquired wisdom that certain conditions or events bring about other conditions or events is an important survival trait. Consider an infant whose first experiences are a jumble of sensations that include hunger, thirst, color, light, heat, cold, and many other stimuli. Gradually, the infant begins to perceive patterns in the jumble and to anticipate connections between actions such as crying and effects such as being fed. Eventually, the infant assembles an inventory of associated perceptions. Along with this growing appreciation for specific causal relations comes the general idea that some events or conditions can be considered causes of other events or conditions.

Thus, our first appreciation of the concept of causation is based on our own observations. These observations typically involve causes with effects that are immediately apparent. For example, changing the position of a light switch on the wall has the instant effect of causing the light to go on or off. There is, however, more to the causal mechanism for getting the light to shine than turning the light switch to the on position. If the electric lines to the building are down because of a storm, turning on the switch will have no effect. If the bulb is burned out, manipulating the switch also will have no effect. One cause of the light going on is having the switch in the proper place, but along with it we must include a supply of power to the circuit, a working bulb, and intact wiring. When all other factors are in place, turning the switch will cause the light to go on, but if one or more of the other factors is not playing its causal role, the light will not go on when the switch is turned. There is a tendency to consider the switch as the unique cause of turning on the light, but we can define a more intricate causal mechanism in which the switch is one component of several. The tendency to identify the switch as the unique cause stems from its usual role as the final factor that acts in the causal mechanism. The wiring can be considered part of the causal mechanism, but after it is installed, it seldom warrants further attention. The switch is typically the only part of the mechanism that needs to be activated to turn on the light. The effect usually occurs immediately after turning the switch, and as a result, we tend to identify the switch as a unique cause. The



- Subjects No. 32, Ministry of Health. London: His Majesty's Stationery Office; 1926.
18. Winkelstein W Jr. Janet Lane-Claypon, a forgotten epidemiologic pioneer. *Epidemiology*. 2006;17:705.
  19. Paneth N, Susser E, Susser M. Origins and early development of the case-control study. Part 2: The case-control study from Lane-Claypon to 1950. *Social Prev Med*. 2002;47:359-365.
  20. Daniel TM. *Wade Hampton Frost, Pioneer Epidemiologist 1880-1938*. Rochester, NY: University of Rochester Press; 2004.

## What Is Causation?

The acquired wisdom that certain conditions or events bring about other conditions or events is an important survival trait. Consider an infant whose first experiences are a jumble of sensations that include hunger, thirst, color, light, heat, cold, and many other stimuli. Gradually, the infant begins to perceive patterns in the jumble and to anticipate connections between actions such as crying and effects such as being fed. Eventually, the infant assembles an inventory of associated perceptions. Along with this growing appreciation for specific causal relations comes the general idea that some events or conditions can be considered causes of other events or conditions.

Thus, our first appreciation of the concept of causation is based on our own observations. These observations typically involve causes with effects that are immediately apparent. For example, changing the position of a light switch on the wall has the instant effect of causing the light to go on or off. There is, however, more to the causal mechanism for getting the light to shine than turning the light switch to the on position. If the electric lines to the building are down because of a storm, turning on the switch will have no effect. If the bulb is burned out, manipulating the switch also will have no effect. One cause of the light going on is having the switch in the proper place, but along with it we must include a supply of power to the circuit, a working bulb, and intact wiring. When all other factors are in place, turning the switch will cause the light to go on, but if one or more of the other factors is not playing its causal role, the light will not go on when the switch is turned. There is a tendency to consider the switch as the unique cause of turning on the light, but we can define a more intricate causal mechanism in which the switch is one component of several. The tendency to identify the switch as the unique cause stems from its usual role as the final factor that acts in the causal mechanism. The wiring can be considered part of the causal mechanism, but after it is installed, it seldom warrants further attention. The switch is typically the only part of the mechanism that needs to be activated to turn on the light. The effect usually occurs immediately after turning the switch, and as a result, we tend to identify the switch as a unique cause. The



inadequacy of this assumption is emphasized when the bulb fails and must be replaced before the light will go on.

## THE CAUSAL PIE MODEL

Causes of disease can be conceptualized in the same way as the causes of turning on a light. A helpful way to think about causal mechanisms for disease is depicted in Figure 3-1.<sup>1</sup> Each *pie* in the diagram represents a theoretical *causal mechanism* for a given disease, sometimes called a *sufficient cause*. The three pies illustrate that there are multiple mechanisms that cause any type of disease. Each individual instance of disease occurs through a single mechanism or sufficient cause. A given causal mechanism requires the joint action of many component factors, or *component causes*. Each component cause is an event or a condition that plays a necessary role in the occurrence of some cases of a given disease. For example, the disease may be cancer of the lung, and in the first mechanism in Figure 3-1, factor C may be cigarette smoking. Other factors include genetic traits or other environmental exposures that play a causal role in cancer of the lung. Some component causes presumably act in many different causal mechanisms. (Terminology note: the *causal pie model* has also been described as the *sufficient-component cause model*.)

### Implications of the Causal Pie Model

#### MULTICAUSALITY

The model of causation shown in Figure 3-1 illuminates several important principles of causation, the most important of which is that every causal mechanism involves the joint action of a multitude of component causes. Consider as an example the cause of a broken hip. Suppose that someone experiences a traumatic injury to the head that leads to a permanent disturbance in equilibrium. Many years later, faulty equilibrium plays a causal role in a fall that occurs while the person is walking on an icy path. The fall results in a broken hip. Other factors

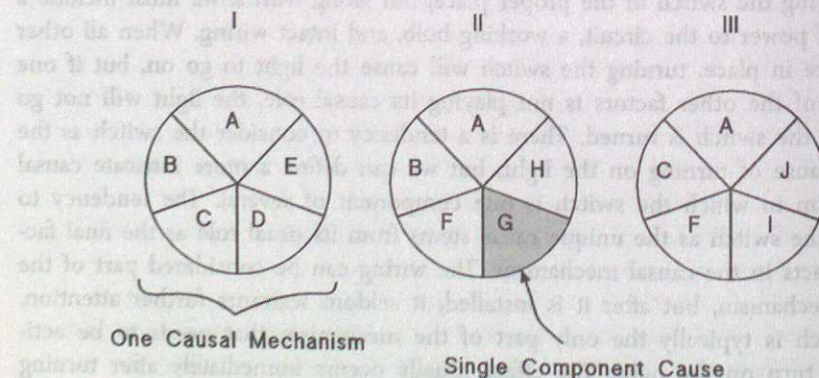


Figure 3-1 Three sufficient causes of a disease.

playing a causal role for the broken hip may include the type of shoe the person was wearing, the lack of a handrail along the path, a sudden gust of wind, and the weight of the person. The complete causal mechanism involves a multitude of factors. Some factors, such as the earlier injury that resulted in the equilibrium disturbance and the weight of the person, reflect earlier events that have had a lingering effect. Some causal components of the broken hip are *genetic*. Genetic factors affect the person's weight, gait, behavior, and recovery from the earlier trauma. Other factors, such as the force of the wind, are *environmental* (nongenetic). There usually are some genetic and some environmental component causes in every causal mechanism. Even an event such as a fall on an icy path that results in a broken hip is part of a complicated causal mechanism that involves many component causes.

### GENETIC VERSUS ENVIRONMENTAL CAUSES

It is a strong assertion that every case of every disease has both genetic and environmental causes. Nevertheless, if all genetic factors that determine disease are taken into account, essentially 100% of disease can be said to be inherited, in the sense that nearly all cases of disease have some genetic component causes. What would be the genetic component causes of someone who gets drunk and is killed in an automobile after colliding with a tree? Genetic traits may lead to psychiatric problems such as alcoholism, which may lead to drunk driving and consequent fatality. It is also possible to claim that essentially 100% of any disease is environmentally caused, even diseases that often are considered to be purely genetic. Phenylketonuria, for example, is considered by many to be purely genetic. Nonetheless, if we consider the disease that phenylketonuria represents to be the mental retardation that may result from it, we can prevent the disease by appropriate dietary intervention. The disease therefore has environmental determinants, and its causes are both environmental and genetic. Although it may seem like an exaggeration to claim that 100% of cases of any disease are environmental and genetic at the same time, it is a good approximation. It may seem counterintuitive, because we cannot manipulate many of the causes in most situations and the ones that can be controlled are usually solely environmental causes, as in the manipulation of diet to prevent the mental retardation of phenylketonuria.

### STRENGTH OF CAUSES

It is common to think that some component causes play a more important role than other factors in the causation of disease. One way this concept is expressed is by the strength of a causal effect. We say that smoking has a strong effect on lung cancer risk because smokers have about 10 times the risk of lung cancer as nonsmokers. We say that smoking has a weaker effect on myocardial infarction because the risk of a heart attack is only about twice as great in smokers as in nonsmokers. With respect to an individual case of disease, however, every component cause that played a role was necessary to the occurrence of that case.



According to the causal pie model, for a given case of disease, there is no such thing as a strong cause or a weak cause. There is only a distinction between factors that were causes and factors that were not causes.

To understand what epidemiologists mean by *strength* of a cause, we need to shift from thinking about an individual case to thinking about the total burden of cases occurring in a population. We can then define a *strong cause* to be a component cause that plays a causal role in a large proportion of cases and a *weak cause* to be a causal component in a small proportion of cases. Because smoking plays a causal role in a high proportion of the lung cancer cases, we call it a strong cause of lung cancer. For a given case of lung cancer, smoking is no more important than any of the other component causes for that case; but on the population level, it is considered a strong cause of lung cancer because it causes such a large proportion of cases.

The strength of a cause defined in this way necessarily depends on the prevalence of other causal factors that produce disease. As a result, the concept of a strong or weak cause cannot be a universally accurate description of any cause. Suppose we say that smoking is a strong cause of lung cancer because it plays a causal role in a large proportion of cases. Exposure to ambient radon gas is considered to be a weaker cause because it has a causal role in a much smaller proportion of lung cancer cases. Imagine that society eventually succeeds in eliminating tobacco smoking, with a consequent reduction in smoking-related cases of lung cancer. One result is that a much larger proportion of the lung cancer cases that continue to occur will be caused by exposure to radon gas; eliminating smoking would strengthen the causal effect of radon gas on lung cancer. This example illustrates that *strength of effect* is not a biologically stable characteristic of a factor. From a biologic perspective, the causal role of a factor in producing disease is neither strong nor weak; the biology of causation corresponds to the identity of the component causes in a causal mechanism and the ways in which they interact to produce disease. The proportion of the population burden of disease that a factor causes, which we use to define the strength of a cause, can change from population to population and over time if there are changes in the distribution of other causes of the disease. The strength of a cause does not portray the biology of causation.

#### INTERACTION BETWEEN CAUSES

The causal pie model posits that several causal components act in concert to produce an effect. *Acting in concert* does not imply that factors must act at the same time. Consider the earlier example of the person who sustained trauma to the head that resulted in an equilibrium disturbance, which led years later to a fall on an icy path. The earlier head trauma played a causal role in the later hip fracture, as did the weather conditions on the day of the fracture. If both factors played a causal role in the hip fracture, they interacted with one another to cause the fracture, despite the fact that their time of action was many years apart. We would say that any and all of the factors in the same causal mechanism interact with one another to cause disease. The head trauma interacted with the weather conditions and with the other component causes, such as the type of footwear, the absence of a handhold, and any other conditions that were necessary to the causal mechanism of the fall and the broken hip that resulted. Each causal pie can

be considered as a set of interacting causal components. This model provides a biologic basis for the concept of interaction that differs from the more traditional statistical view of interaction. The implication of this difference is discussed in Chapter 11.

#### SUM OF ATTRIBUTABLE FRACTIONS

Consider the data in Table 3-1, which shows the rates of head and neck cancer according to smoking status and alcohol exposure. Suppose that the differences in the rates reflect causal effects, so that confounding can be ignored. Among those who are smokers and alcohol drinkers, what proportion of the cases of head and neck cancer that occur is attributable to the effect of smoking? We know that the rate for these people is 12 cases per 10,000 person-years. If these same people were not smokers, we can infer that their rate of head and neck cancer would be 3 cases per 10,000 person-years. If this difference reflects the causal role of smoking, we can infer that 9 of every 12 cases (75%) are attributable to smoking among those who smoke and drink alcohol. If we turn the question around and ask what proportion of disease among these same people is attributable to alcohol drinking, we would be able to attribute 8 of every 12 cases (67%) to alcohol drinking.

Can we attribute 75% of the cases to smoking and 67% to alcohol drinking among those who are exposed to both? The answer is yes, because some cases are counted more than once as a result of the interaction between smoking and alcohol consumption. These cases are attributable to both smoking and alcohol drinking because both factors played a causal role in producing them. One consequence of interaction is that the proportions of disease attributable to various component causes do not sum to 100%.

A widely discussed but unpublished paper from the 1970s written by scientists at the National Institutes of Health proposed that as much as 40% of cancer is attributable to occupational exposures. Many scientists thought that this fraction was an overestimate and argued against this claim.<sup>2,3</sup> One of the arguments used in rebuttal was as follows:  $x$  percent of cancer is caused by smoking,  $y$  percent by diet,  $z$  percent by alcohol, and so on; when all of these percentages are summed, only a small percentage, much less than 40%, is left for occupational causes. This rebuttal, however, is fallacious because it is based on the naive view that every case of disease has a single cause and that two causes cannot both contribute to the same case of cancer. Because diet, smoking, asbestos, and various occupational exposures and other factors interact with one another and with genetic

Table 3-1 HYPOTHETICAL RATES  
OF HEAD AND NECK CANCER  
(CASES PER 10,000 PERSON-YEARS)  
ACCORDING TO SMOKING STATUS  
AND ALCOHOL DRINKING

Smoking Status	Alcohol Drinking	
	No	Yes
Nonsmoker	1	3
Smoker	4	12



factors to cause cancer, each case of cancer can be attributed repeatedly to many separate component causes. The sum of disease attributable to various component causes has no upper limit.

#### INDUCTION TIME

Because the component causes in a given causal mechanism do not act simultaneously, there usually is a period of time between the action of a component cause and the completion of a sufficient cause. The only exception is the last component cause to act in a given causal mechanism. The last-acting component cause completes the causal mechanism, and we can say that disease begins concurrently with its action. For earlier-acting component causes, we can define the *induction period* as the time interval that begins concurrently with the action of a component cause and ends when the final component cause acts and the disease occurs. For example, in the illustration of the fractured hip, the induction time between the head trauma that resulted in an equilibrium disturbance and the later hip fracture was many years. The induction time between the decision to wear nongripping shoes and the hip fracture might have been a matter of minutes or hours. The induction time between the gust of wind that triggered the fall and the hip fracture might have been seconds or less.

In an individual instance, we usually cannot know the exact length of an induction period, because we cannot be sure of the causal mechanism that produces disease in an individual instance nor when all the relevant component causes in that mechanism exerted their causal action. With research data, however, we can learn enough to characterize the induction period that relates the action of a single component cause to the occurrence of disease in general. An example of a lengthy induction time is the cause-effect relation between exposure of a female fetus to diethylstilbestrol (DES) and her subsequent development of adenocarcinoma of the vagina. The cancer generally occurs after the age of 15 years. Because the causal exposure to DES occurs during gestation, there is an induction time of more than 15 years for carcinogenesis. During this time, other causes presumably operate; some evidence suggests that hormonal action during adolescence may be part of the mechanism.<sup>4</sup>

The causal pie model makes it clear that it is incorrect to characterize a disease itself as having a lengthy or brief induction time. The induction time can be conceptualized only in relation to a specific component cause. We can say that the induction time relating DES to clear cell carcinoma of the vagina is at least 15 years, but we cannot say that 15 years is the minimum 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 always 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 than that of DES. Induction time characterizes a specific cause-effect pair rather than only the effect.

In carcinogenesis, the terms *initiator* and *promoter* are used to refer to 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, but this characterization is a misconception. Any late-acting component

in the causal process, such as a promoter, will have a short induction time, and the induction time will always be zero for the last component cause (eg, the gust of wind causing the broken hip in the earlier example), because after the final causal component acts, disease has occurred. At that point, however, the presence of disease is not necessarily apparent. A broken hip may be apparent immediately, but a cancer that has just been caused may not become noticed or diagnosed for an appreciable time. The time interval between disease occurrence and its subsequent detection, whether by medical testing or by the emergence of symptoms, is called the *latent period*.<sup>4</sup> The length of the latent period can be reduced by improved methods of disease detection. The induction period, however, cannot be reduced by early detection of disease, because there is no disease to detect until after the induction period is over. Practically, it may be difficult to distinguish between the induction period and the latent period, because there may be no way to establish when the disease process began if it is not detected until later. Diseases such as slow-growing cancers may appear to have long induction periods with respect to many causes, in part because they have long latent periods.

Although it is not possible to reduce the induction period 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 focusing on causes that are more proximal to the disease occurrence. Biomarkers may reflect the effects on the organism of agents that have acted at an earlier time.

#### IS A CATALYST A CAUSE?

Some agents may have a causal action by shortening the induction time of other agents. Suppose that exposure to factor A leads to epilepsy after an average interval of 10 years. It may be that exposure to drug B can shorten this interval to 2 years. Is B acting as a catalyst or as a cause of epilepsy? The answer is both; a catalyst is a cause. Without B, the occurrence of epilepsy comes 8 years later than it comes with B, so we can say that B causes the epilepsy to occur earlier. It is not sufficient to argue that the epilepsy would have occurred anyway and therefore that B is not a cause of its occurrence. First, it would not have occurred at that time, and the time of occurrence is considered part of the definition of an event. Second, epilepsy will occur later only if the person survives an additional 8 years, which is not certain. Agent B therefore determines when the epilepsy occurs, and it can determine whether it occurs at all. For this reason, we consider any agent that acts as a catalyst of a causal mechanism, shortening the induction period for other agents, to be a cause. Similarly, any agent that postpones the onset of an event, drawing out the induction period for another agent, we consider to be a preventive. It should not be too surprising to equate postponement with prevention; we routinely use such an equation when we employ the euphemism that we prevent death, which can only be postponed. We prevent death at a given time in favor of death at a later time. Similarly, slowing the process of atherosclerosis can result in postponement (and thereby prevention) of cardiovascular disease and death.



## THE PROCESS OF SCIENTIFIC INFERENCE

Much epidemiologic research is aimed at uncovering the causes of disease. Now that we have a conceptual model for causes, how do we determine whether a given relation is causal? Some scientists refer to checklists for causal inference, and others focus on complicated statistical approaches, but the answer to this question is not to be found in checklists or statistical methods. The question itself is tantamount to asking how we apply the scientific method to epidemiologic research. This question leads directly to the philosophy of science, a topic that goes well beyond the scope of this book. Nevertheless, it is worthwhile to summarize two of the major philosophic doctrines that have influenced modern science.

## Induction

Since the rise of modern science in the 17th century, scientists and philosophers have puzzled over the question of how to determine the truth about assertions that deal with the empirical world. From the time of the ancient Greeks, deductive methods have been used to prove the validity of mathematic propositions. These methods enable us to draw airtight conclusions because they are self-contained, starting with a limited set of definitions and axioms and applying rules of logic that guarantee the validity of the method. Empirical science is different, however. Assertions about the real world do not start from arbitrary axioms, and they involve observations on nature that are fallible and incomplete. These stark differences from deductive logic led early modern empiricists, such as Francis Bacon, to promote what they considered a new type of logic, which they called *induction* (not to be confused with the concept of an induction period). *Induction* was an indirect method used to gain insight into what has been metaphorically described as the fabric of nature.

The method of induction starts with observations on nature. To the extent that the observations fall into a pattern, they are said to induce in the mind of the observer a suggestion of a more general statement about nature. The general statement can range from a simple hypothesis to a more profound natural law or natural relation. The statement about nature is reinforced with further observations or refuted by contradictory observations. For example, suppose an investigator in New York conducts an experiment to determine the boiling point of water and observes that the water boils at 100°C. The experiment is repeated many times, each time showing that the water boils at about 100°C. By induction, the investigator concludes that the boiling point of water is 100°C. The induction itself involves an inference beyond the observations to a general statement that describes the nature of boiling water. As induction became popular, it was seen to differ considerably from deduction. Although not as well understood as deduction, the approach was considered a new type of logic, *inductive logic*.

Although induction, with its emphasis on observation, represented an important advance over the appeal to faith and authority that characterized medieval scholasticism, it was not long before the validity of the new logic was questioned. The sharpest criticism came from the skeptical philosopher David Hume, who pointed out that induction had no logical force. Rather, it amounted to the

assumption that what had been observed in the past would continue to occur in the future. When supporters of induction argued that induction was a valid process because it had been seen to work on numerous occasions, Hume countered that the argument was an example of circular reasoning that relied on induction to justify itself. Hume was so profoundly skeptical that he distrusted any inference based on observation because observations depend on sense perceptions and are therefore subject to error.

## Refutationism

Hume's criticisms of induction have been a powerful force in modern scientific philosophy. The most influential reply to Hume was offered by Karl Popper. Popper accepted Hume's point that in empirical science one cannot prove the validity of a statement about nature in any way that is comparable with a deductive proof. Popper's philosophy, known as *refutationism*, held that statements about nature can be "corroborated" by evidence, but corroboration does not amount to a logical proof. On the other hand, Popper also asserted that statements about nature can be refuted by deductive logic. To grasp the point, consider the earlier example of observing the boiling point of water. The refutationist view is that the repeated experiments showing that water boils at 100°C corroborate the hypothesis that water boils at this temperature, but they do not prove it.<sup>5</sup> A colleague of the New York researcher who works in Denver, a city located at high altitude, would find that water there boils at 94°C. This single contrary observation carries more weight regarding the hypothesis about the boiling point of water than thousands of repetitions of the initial experiment at sea level.

The asymmetric implications of a refuting observation compared with supporting observations are the essence of the refutationist view. This school of thought encourages scientists to subject a new hypothesis to rigorous tests that may falsify the hypothesis in preference to repetitions of the initial observations that add little beyond the weak corroboration that replication can supply. The implication for the method of science is that hypotheses should be evaluated by subjecting them to crucial tests. If a test refutes a hypothesis, a new hypothesis needs to be formulated that can then be subjected to further tests. After finding that water boils in Denver at a lower temperature than it boils in New York, the investigator must discard the hypothesis that water boils at 100°C and replace it with a more refined hypothesis, such as one that will explain the difference in boiling points under different atmospheric pressures. This process describes an endless cycle of *conjecture and refutation*. The conjecture, or hypothesis, is the product of scientific insight and imagination. It requires little justification except that it can account for existing observations. A useful approach is to pose competing hypotheses to explain existing observations and to test them against one another. The refutationist philosophy postulates that all scientific knowledge is tentative because it may one day need to be refined or even discarded. In this philosophy, what we call scientific knowledge is a body of currently unrefuted hypotheses that appear to explain existing observations.

How can an epidemiologist apply refutationist thinking to his or her work? If causal mechanisms are stated specifically, an epidemiologist can construct crucial



tests of competing hypotheses. For example, when toxic shock syndrome was first studied, there were two competing hypotheses about the origin of the toxin. In one, the toxin responsible for the disease was a chemical in the tampon, and women using tampons were exposed to the toxin directly from the tampon. In the other hypothesis, the tampon acted as a culture medium for staphylococci that produced the toxin. Both hypotheses explained the correlation of toxic shock occurrence and tampon use. The two hypotheses, however, led to opposite predictions about the relation between the frequency of changing tampons and the risk of toxic shock. If chemical intoxication were the cause, more frequent tampon changes would lead to more exposure to the toxin and possible absorption of a greater overall dose. This hypothesis predicted that women who changed tampons more frequently would have a higher risk of toxic shock syndrome than women who changed tampons infrequently. The culture-medium hypothesis predicted that the women who change tampons frequently would have a lower risk than those who left the tampon in for longer periods, because a short duration of use for each tampon would prevent the staphylococci from multiplying enough to produce a damaging dose of toxin. Epidemiologic research, which showed that infrequent changing of tampons was associated with greater risk of toxic shock, refuted the chemical theory.

Critics of refutationism point out that refutation is not logically certain because it depends on theories, assumptions, and observations, all of which are susceptible to error. In epidemiology, for example, any study result may be influenced by an obscure bias, which is an inescapable source of uncertainty. Among the dissenting philosophic views is that of Thomas Kuhn,<sup>6</sup> who held that it is ultimately the collective beliefs of the community of scientists that determines what is accepted as truth about nature. According to Kuhn, the truth is not necessarily objective but rather something determined by consensus. Feyerabend,<sup>7</sup> another skeptic, held that science proceeds through intellectual anarchy, without any coherent method. A more moderate although still critical view was taken by Haack.<sup>8,9</sup> She saw science as an extension of everyday inquiry, employing pragmatic methods that she likened to solving a crossword puzzle, integrating clues with other answers in a trial-and-error approach. Despite these criticisms, refutationism has been a positive force in science by encouraging bold, testable theories and then fostering a valuable skeptical outlook by subjecting those theories to rigorous challenges.

### Causal Criteria

Earlier we said that there is no simple checklist that can determine whether an observed relation is causal. Nevertheless, attempts at such checklists have appeared. Most of these lists stem from the canons of inference described by John Stuart Mill.<sup>10</sup> The most widely cited list of causal criteria, originally posed as a list of standards, is attributed to Hill,<sup>11</sup> who adapted them from the U.S. Surgeon General's 1964 report on Smoking and Health.<sup>12</sup> The Hill standards, often labeled the Hill criteria, are listed in Table 3-2, along with some problems related to each of the criteria.

Although Hill did not propose these criteria as a checklist for evaluating whether a reported association could be interpreted as causal, many others have

Table 3-2 CAUSAL CRITERIA OF HILL

Criterion	Problems with the Criterion
1. Strength	Strength depends on the prevalence of other causes; it is not a biologic characteristic and can be confounded.
2. Consistency	Causal relations have exceptions that are understood best with hindsight.
3. Specificity	A cause can have many effects.
4. Temporality	It may be difficult to establish the temporal sequence between cause and effect.
5. Biologic gradient	It can be confounded; threshold phenomena would not show a progressive relation.
6. Plausibility	Too subjective
7. Coherence	How does it differ from consistency or plausibility?
8. Experimental evidence	Not always available
9. Analogy	Analogies abound.

attempted to apply them in that way. Admittedly, the process of causal inference as described earlier is difficult and uncertain, making the appeal of a simple checklist undeniable. Unfortunately, this checklist, like all others with the same goal, fails to deliver on the hope of clearly distinguishing causal from noncausal relations. Consider the first criterion, strength. It is tempting to believe that strong associations are more likely to be causal than weak ones, but as we saw in our discussion of causal pies, not every component cause has a strong association with the disease that it produces; strength of association depends on the prevalence of other factors. Some causal associations, such as the association between cigarette smoking and coronary heart disease, are weak. Furthermore, a strong association can be noncausal, a confounded result stemming from the effect of another risk factor for the disease that is highly correlated with the one under study. For example, birth order is strongly associated with the occurrence of Down syndrome, but it is a confounded association that is completely explained by the effect of maternal age. If weak associations can be causal and strong associations can be noncausal, it does not appear that strength of association can be considered a criterion for causality.

The third criterion (see Table 3-2), specificity, suggests that a relation is more likely to be causal if the exposure is related to a single outcome rather than myriad outcomes. This criterion is misleading because it implies, for example, that the more diseases with which smoking is associated, the greater the evidence that smoking is not causally associated with any of them. The fifth criterion, biologic gradient, is often taken as a sign of a causal relation, but it can just as well result from confounding or other biases as from a causal connection. The relation between Down syndrome and birth order, mentioned earlier, shows a biologic gradient despite being completely explained by confounding from maternal age.

Other criteria from Hill's list are vague (eg, consistency, plausibility, coherence, analogy) or do not apply in many settings (eg, experimental evidence). The only characteristic on the list that is truly a causal criterion is temporality, which implies that the cause comes before the effect. This criterion, which is part of



the definition of a cause, is a useful one, although it may be difficult to establish the proper time sequence for cause and effect. For example, does stress lead to overeating, or does overeating lead to stress? It usually is better to avoid a checklist approach to causal inference and instead consider approaches such as conjecture and refutation. Checklists lend a deceptive kind of mindless authority to an inherently imperfect and creative process. In contrast, causal inference based on conjecture and refutation fosters a highly desirable critical scrutiny.

Although checklists may not be appropriate for causal inference, the points laid out by Hill are still important considerations. The criteria may be useful when applied in the context of specific hypotheses. For example, Weiss observed that the specificity of effects might be important in inferring the beneficial effect of sigmoidoscopy in screening for colorectal cancer if the association between sigmoidoscopy and reduced death from colorectal cancer is stronger for cancer occurring at sites within reach of a sigmoidoscope.<sup>13</sup>

### Generalization in Epidemiology

A useful way to think of scientific generalization is to consider a generalization to be the elaboration of a scientific theory. A given study may test the viability of one or more theories. Theories that survive such tests can be viewed as general statements about nature that tell us what to expect in people or settings that were not studied. Because theories can be incorrect, scientific generalization is not a perfect process. Formulating a theory is not a mathematical or statistical process, and generalization should not be considered a statistical exercise. It is the process of causal inference itself.

Many people believe that generalizing from an epidemiologic study involves a mechanical process of making an inference about a target population of which the study population is considered a sample. This type of generalization does exist, in the field of survey sampling. In survey sampling, researchers draw samples from a population to avoid the expense of studying the entire population, which makes the statistical representativeness of the sample the main concern for generalizing to the source population.

Although survey sampling is an important tool for characterizing a population efficiently and may be used in some epidemiologic applications, such as prevalence surveys, it is a mechanical tool that does not always share the same goals as science. Survey sampling is useful for problems such as trying to predict how a population will vote in an election or what type of laundry soap the people in a region prefer. These are characteristics that depend on attitudes and for which there is little coherent biologic theory on which to base a scientific generalization. Survey results may be quickly outdated (eg, election polls may be repeated weekly or even daily) and do not apply outside the populations from which the surveys were conducted. (Disclaimer: I am not saying that social science is not science or that we cannot develop theories about social behavior. I am saying only that surveys about the current attitudes of a specific group of people are not the same as social theories.) Even if survey sampling is used to characterize the prevalence of disease or the medical needs of a population, the objectives are pragmatic rather than scientific and may not apply outside the study population. Scientific results

from epidemiologic studies, in contrast, seldom need to be repeated weekly to see if they still apply. An epidemiologic study conducted in Chicago showing that exposure to ionizing radiation causes cancer does not need to be repeated in Houston to determine whether ionizing radiation also causes cancer in people living in Houston. Generalization about ionizing radiation and cancer is based on understanding of the underlying biology rather than on statistical sampling.

It may be helpful to consider the problem of scientific generalization about causes of cancer from the point of view of a biologist studying carcinogenesis in mice. Most researchers who study cancer in animals do so because they would like to understand better the causes of human cancer. If scientific generalization depended on having studied a statistically representative sample of the target population, researchers studying mice would have nothing to contribute to the understanding of human cancer. Mouse researchers obviously do not study representative samples of people; they do not even study representative samples of mice. Instead, they seek mice that have uniformly similar genes and perhaps certain biologic characteristics. In choosing mice to study, they have to consider mundane issues such as the cost of the mice. Although researchers studying animals are unlikely to worry about whether their mouse or rabbit subjects are statistically representative of all mice or rabbits, they may consider whether the biology of the animal population they are studying is similar to (and representative of) that of humans. This type of representativeness, however, is not statistical representativeness based on sampling from a source population; it is a biologic representativeness based on scientific knowledge. Despite the absence of statistical representativeness, no one seriously doubts the contribution that animal research can make to the understanding of human disease.

Many epidemiologic activities, such as measuring the prevalence of patients in need of dialysis, do require surveys to characterize a specific population, but these activities are usually examples of applied epidemiology rather than the science of epidemiology. The activities of applied epidemiology involve taking already established epidemiologic knowledge and applying it to specific settings, such as preventing malaria transmission by reducing the mosquito vector population or reducing lung cancer and cardiovascular disease occurrence by implementing an antismoking campaign. The activities of epidemiologic research, as in laboratory science, move away from the specific toward the general. We make specific observations in research studies and then hope to generalize from them to a broader base of understanding. This process is based more on scientific knowledge, insight, and conjecture about nature than it is on the statistical representativeness of the actual study participants. This principle has important implications for the design and interpretation of epidemiologic studies (see Chapter 7).

### QUESTIONS

1. Criticize the following statement: The cause of tuberculosis is infection with the tubercle bacillus.
2. A trait in chickens called yellow shank occurs when a specific genetic strain of chickens is fed yellow corn. Farmers who own only this strain of chickens



observe the trait to depend entirely on the nature of the diet, specifically whether they feed their chickens yellow corn. Farmers who feed all of their chickens only yellow corn but own several strains of chicken observe the trait to be genetic. What argument could you use to explain to both kinds of farmer that the trait is both environmental and genetic?

3. A newspaper article proclaims that diabetes is neither genetic nor environmental but multicausal. Another article announces that one half of all colon cancer cases are linked to genetic factors. Criticize both messages.

4. Suppose a new treatment for a fatal disease defers the average time before onset of death among those with the disease for 20 years beyond the time when they would have otherwise died. Is it proper to say that this new treatment reduces the risk of death, or does it merely postpone death?

5. It is typically more difficult to study an exposure-disease relation that has a long induction period than one that has a short induction period. What difficulties ensue because the exposure-disease induction period is long?

6. Suppose that both A and B are causes of a disease that is always fatal, so that the disease can occur only once in a single person. Among people exposed to both A and B, what is the maximum proportion of disease that can be attributed to either A or B? What is the maximum for the sum of the amount attributable to A and the amount attributable to B? Suppose that A and B exert their causal influence only in different causal mechanisms, so that they never act through the same mechanism. Would that change your answer?

7. Adherents of induction claim that we all use this method of inference every day. We assume, for example, that the sun will rise tomorrow as it has in the past. Critics of induction claim that this knowledge is based on belief and assumption and that it is no more than a psychological crutch. Why should it matter to a scientist whether scientific reasoning is based on induction or on a different approach, such as conjecture and refutation?

8. Give an example of competing hypotheses for which an epidemiologic study would provide a refutation of at least one.

9. Could a causal association fail to show evidence of a biologic gradient (ie, Hill's fifth criterion)? Explain.

10. Suppose you are studying the influence of socioeconomic factors on cardiovascular disease. Would the study be more informative if (1) the study participants had the same distribution of socioeconomic factors as the general population or (2) the study participants were recruited so that there were equal numbers of participants in each category of the socioeconomic variables? Why?

## REFERENCES

1. Rothman KJ. Causes. *Am J Epidemiol.* 1976;104:587-592.
2. Higginson J. Proportion of cancer due to occupation. *Prev Med.* 1980;9:180-188.
3. Ephron E. *The Apocalypics. Cancer and the Big Lie.* New York: Simon & Schuster, 1984.
4. Rothman KJ. Induction and latent period. *Am J Epidemiol.* 1981;114:253-259.
5. Magee B. *Philosophy and the Real World. An Introduction to Karl Popper.* La Salle, Illinois, Open Court, 1985.
6. Kuhn T. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press, 1962.
7. Feyerabend P. *Against Method.* New York: New Left Books, 1975.
8. Haack S. *Defending Science Within Reason.* Amherst: Prometheus Books, 2003.
9. Haack S. Trial and error: the Supreme Court's philosophy of science. *Am J Public Health.* 2005;95(suppl 1):S66-S73.
10. Mill JS. *A System of Logic, Ratiocinative and Inductive.* 5th ed. London: Parker, Son & Bowin, 1862.
11. Hill AB. The environment and disease: Association or causation? *Proc R Soc Med.* 1965;58:295-300.
12. U.S. Department of Health, Education and Welfare. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service.* Public Health Service Publication No. 1103. Washington, DC: Government Printing Office, 1964.
13. Weiss NS. Can the "specificity" of an association be rehabilitated as a basis for supporting a causal hypothesis? *Epidemiology.* 2002;13:6-8.