The prevention of disease has long been based implicitly on taking action on the assumption that a disease is caused by a factor that can be controlled. Early examples include the experimental evidence generated by Lind, showing that consumption of oranges and lemons prevented scurvy, and Snow's observations on cholera occurrence in London, showing a disease pattern consistent with water-borne transmission (Rosen, 1993). In these examples, preventive steps followed: After Lind's experiment, the diets of the British navy were supplemented with citrus fruits; and after Snow's observational study, steps were taken to ensure that the source of water was changed in the affected areas of London. Over the ensuing centuries, infectious agents were causally linked to specific diseases, and prevention was accomplished by interrupting transmission and by vaccines. During the twentieth century, public health was threatened by parallel epidemics of chronic diseases, including cancer; and as the causal agents were identified, a broad range of preventive strategies were implemented.

The concept of causation has long had a central role in the application of epidemiologic evidence for controlling cancer. The designation of a risk factor as "causal" has been the starting point for initiating cancer prevention programs based on reducing exposure to the risk factor. Although the concept of causation itself remains a matter of continuing discussion among philosophers and others, use of the term in public health implies that the evidence supporting causality of association has reached a critical threshold of certainty and that reduced exposure can be expected to follow by reduced disease occurrence. Over the last 50 years, identification of the causes of cancer has been the primary focus of most epidemiologic research on cancer; only recently has attention shifted toward identifying genetic determinants of susceptibility and markers of the early stages of carcinogenesis.

There are numerous examples of how identifying a cause of cancer has led to intervention and reduction of cancer occurrence. Tobacco use and cancer of the lung is a notable example for its historical precedent and for the framework applied to the scientific evidence as the causality of the association was evaluated (US Department of Health Education and Welfare—DHEW, 1964; White, 1990). The range of causal risk factors for cancer is broad, including infectious agents (e.g., human papillomavirus and cervical cancer), physical agents (e.g., ionizing radiation and leukemia), inhaled agents (e.g., radon and lung cancer), pharmaceutical agents and hormones (e.g., diethylstilbestrol and adenocarcinoma of the vagina), food contaminants (e.g., aflatoxin and liver cancer), workplace exposures (e.g., asbestos), life style-related exposures (e.g., alcohol consumption), and genetic mutations (e.g., Li-Fraumeni syndrome). These and other factors considered to be causes of cancer have been given this label only after the accumulation of sufficient evidence, in most instances derived from both epidemiologic and laboratory research.

This chapter provides an overview of causal inference with a focus on the interpretation of epidemiologic data on cancer risk. It begins with an introduction to the centuries-old discussion on cause and causation and next considers the epidemiologic concept of causation, setting the discussion in the context of current understanding of carcinogenesis as a multistep process. The criteria for causation, often attributed to the British medical statistician Sir Austin Bradford Hill (Hill, 1965) or to the 1964 Report of the U.S. Surgeon General on tobacco (US Department of Health Education and Welfare—DHEW, 1964), have provided a framework for evaluating evidence to judge the causality of associations.

These criteria are addressed in depth, and their application is illustrated with the example of smoking, both active and passive, and lung cancer. The chapter concludes with a consideration of emerging issues concerned with causation, including the interpretation of data coming from the new technologies of contemporary "molecular epidemiology" and new approaches to evaluating causation.

CONCEPTS OF CAUSATION

At its foundation, "cause" is not knowable with certainty. This fact underlies much of the methodologic and conceptual confusion that often swirls around claims of causation based on scientific data. The fundamental intuition underlying the causal concept is that event "A" somehow produces another event, "B." However, the "production" of "B" is not observable. The philosopher Bryan Magee summarized this conundrum eloquently.

It seems to be impossible for us to form any conception of an ordered world at all without the idea of there being causal connections between events. But when we pursue this idea seriously we find that causal connection is not anything we ever actually observe, nor ever can observe. We may say that Event A causesEvent B, but when we examine the situation we find that what we actually observed is Event A followed by Event B. There is not some third entity between them, a casual link, which we also observe. . . . So we have this indispensable notion of cause at the very heart of our conception of the world, and of our understanding of our own experience, which we find ourselves quite unable to validate by observation or experience . . . It actually purports to tell us how specific material events are related to each other in the real world, yet it is not derived from, nor can it be validated by, observation of that world. This is deeply mysterious. (Magee, 2001)

The fact that causation is not directly observable means that scientists and philosophers have had to develop a set of constructs and heuristics by which to define a "cause" operationally. These constructs typically have two components: a predictive or associational one, determined empirically, and an explanatory one, based on a proposed underlying mechanism. All causal claims rest on these twin pillars; an association with no plausible mechanistic basis is typically not accepted as causal, and a proposed mechanism, however well founded, cannot be accepted as the basis for a causal claim without empirical demonstration that the effect occurs more often in the presence of the purported cause than in its absence. However, these components need not contribute equally, and various causal claims may rest on quite different balances of contributions of empirical and mechanistic information.

Underlying any operational definition of causality must be an ontologic one: that is, how a cause is defined in principle. A particularly useful, widely accepted definition in both philosophy and epidemiology is the "counterfactual" notion of causation. This concept had its origins at least as far back as the English philosopher David Hume (1711–1776) (Hume, 1739). During the twentieth century, this concept was further developed and applied by statisticians, philosophers, and epidemiologists (Bunge, 1959; Lewis, 1973; Rubin, 1974; Robins, 1986, 1987; Greenland, 1990; Neyman, 1990; Greenland et al., 1999;
The counterfactual definition holds that something is a cause of a given outcome if, when the same individual is observed with and without a purported cause and without changing any other characteristic of that individual, a different outcome would be observed. For example, the counterfactual state for a smoker is the same individual never having smoked. The state that cannot be observed is called the counterfactual state: literally, counter to the observed facts. The impossibility of observing the counterfactual state is what makes all causal claims subject to uncertainty.

The above definition is deterministic; that is, the outcome always occurs in the presence of the cause and never occurs without it. However, health research rarely deals with either a cause that inevitably produces certain outcomes, or outcomes that cannot occur absent specific causes. For example, smokers do not always get lung cancer, and never-smokers do develop this malignancy. Therefore, the counterfactual definition must be expanded to encompass the notion of a probabilistic outcome. That is, the formal definition of a cause in epidemiology requires that a factor X be associated with a difference in the probability of an outcome. For example, if X may take on two different values, y or z:

**Condition 1:** observed association

\[
\Pr(\text{outcome} \mid X = y) \neq \Pr(\text{outcome} \mid X = z)
\]

Properly designed studies provide a scientific basis for inferring what the outcome of the counterfactual state would be and permit the related uncertainty to be quantified. In a laboratory, scientists are able to predict the outcome in this counterfactual state, generally with a high degree of confidence, by repeating an experimental procedure with every factor tightly controlled, varying only the factor of interest. In observational studies of humans, however, researchers must try to infer what the outcome would be in a counterfactual state by studying another group of persons who, at least on average, are substantively different from the exposed group in only one variable: the exposure under study. The outcome of this second group is used to represent what would have occurred in the original group if it were observed with an exposure different from that which actually existed (Greenland, 1990). In the case of smoking and disease, this comparison is between disease risk in smokers and nonsmokers.

Simply observing a difference in the probability of an outcome between two groups that differ on X is not sufficient condition for causation because it does not distinguish between causation and spurious or indirect association, produced by "confounders," or ancillary causes. The notion of "causation" requires that the cause somehow actively "produce" its effect, which is captured operationally by the requirement that active manipulation of the cause should produce a change in the probability of the outcome. For example, if one saw that students with poor visual acuity typically sat closer to the front of a classroom, one would not call the seating arrangement a "cause" of their poor eyesight unless it could be shown that seating them farther back improved it. The notation that captures this idea is one that introduces an operator, not part of traditional statistical notation "Set \(X = x\)," which corresponds to actively setting a risk factor X equal to some value x, rather than simply observing that the factor is equal to x. Thus the counterfactual notion of probabilistic causation for a risk factor X requires condition 2.

**Condition 2:** no confounding

\[
\Pr(\text{outcome} \mid \text{Set}(X = x)) = \Pr(\text{outcome} \mid X = x)
\]

If we put together condition 1—that there is an observed association between cause and effect—with condition 2—that there is no other indirect cause responsible for the observed effect—we have the counterfactual condition for probabilistic causation, expressed as follows.

**Condition 1 + Condition 2 = Causality condition**

\[
\Pr(\text{outcome} \mid \text{Set}(X = y)) \neq \Pr(\text{outcome} \mid \text{Set}(X = z))
\]

This condition states that if the probability of an outcome changes when risk factor X is actively changed from z to y, then X is regarded as a cause of the outcome.

In the randomized controlled trial, a risk factor is actively manipulated. Understanding the role of randomization can deepen insights into the interpretation of nonrandomized designs used in epidemiology. Randomization has two critical consequences: (1) it makes exposure to a proposed causal factor independent of potentially confounding factors; and (2) it provides a known probability distribution for the potential outcomes in each group under a given mathematical hypothesis (i.e., the null) (Greenland, 1990). Randomization does not necessarily free the inference from an individual randomized study from unmeasured confounding (it does so only on average). Randomization does imply that measures of uncertainty about causal estimates from randomized studies have an experimental foundation. In the absence of randomization, uncertainty about causal effects depends in part on the confidence that all substantive confounding has been eliminated or controlled by either the study design or the analysis. The level of confidence is ultimately based on scientific judgment and consequently is subject to uncertainty and questioning.

One way to increase that confidence is to repeat the study. Similar results in a series of randomized studies make it increasingly unlikely that unmeasured confounding is accounting for the findings, as the process of randomization makes the mathematical probability of such confounding progressively smaller as the sample size or number of studies increase. In observational studies, however, increasing the number of studies may reduce the random component of uncertainty, but not necessarily the systematic component attributable to confounding. Without randomization, there is no mathematical basis for assuming that an imbalance of unknown confounders decreases with an increase in the number of studies. However, if observational studies are repeated in different settings with different persons, different eligibility criteria, and/or different exposure opportunities, each of which might eliminate another source of confounding from consideration, the confidence that unmeasured confounders are not producing the findings is increased. How many studies need to be done, how diverse they need to be, and how relevant they are to the question at hand are matters of scientific judgment, and explicit criteria cannot be offered.

Confidence that unmeasured confounding is not producing the observed results is further increased by understanding the biologic process by which the exposure might affect the outcome. This understanding allows better identification and measurement of relevant confounders, making it more unlikely that unmeasured factors are of concern. Biologic understanding can also serve as the basis for a judgment that the observed difference in outcome frequency could be produced only by an implausible degree of confounder imbalance between exposed and unexposed groups. Thus, causal conclusions from observational studies typically require more and stronger biologic evidence to support plausibility and to exclude confounding than is needed for causal inferences based on randomized studies.

**COMPONENT CAUSE MODEL**

Causes defined in the manner described above can be viewed as working together in many ways. In 1976 Rothman proposed a useful framework for considering multiple-cause diseases (Rothman, 1976) that has ready extension to the causation of cancer, as most cancers have several causes. Rothman proposed that a disease may have several sufficient causes, each accounting for some proportion of the cases in the population, and that each cause may have several components (Fig. 1-1) (Rothman, 1976). With this model, each component of the three complete causes must be present for disease to develop. For example, sufficient cause I would be incomplete if A were not present; and because A is a component of each of the three complete causes, it is a necessary cause.

Rothman's model is useful for considering causation of malignancy, particularly for most of the cancers for which multiple genetic and environmental risk factors may play a role in a multistage process that transforms a normal cell into a malignant cell. There may be multiple ways to complete this sequence of changes, involving the actions of different complexes of factors, analogous to complete causes I, II, and III in Figure 1-1. The sufficient causes might include multiple envi-
environmental and genetic risk factors, including environmental exposures and genes determining carcinogen metabolism and DNA repair. Individual cases would result from having the full complement of components for a complete cause. We know, for example, that cigarette smoking is a powerful cause of lung cancer, but not all smokers develop lung cancer, implying that this factor may need to act in combination with other factors, perhaps genetic, to complete one sufficient cause for lung cancer. Some sufficient causes for lung cancer do not include smoking, as some percentage of lung cancer cases occur in never-smokers (about 5%-10% of cases in the United States at present) (Alberg and Samet, 2003).

Rothman’s model has one significant implication for considering the burden of cancer attributable to a particular risk factor, a calculation often made when assigning priorities to prevention initiatives. The presence of several components in the same complete causes (e.g., A and B in causes I and II) implies that the cases associated with these causes might be prevented by eliminating exposure to either A or B. The burden of disease to be prevented exceeds 100%, as the attributable risk estimates for A and B would include some of the same cases. In some past reviews of the burden of preventable cancer, the assumption was made incorrectly that the attributable risks associated with various causal factors should add up to 100% (Samet and Lerchen, 1984).

**CRITERIA FOR CAUSALITY**

Epidemiologists and other public health researchers have needed pragmatic definitions of causation to support the translation of research evidence into interventions directed at reducing the exposure to causal risk factors (Susser, 1973, 1991; Olsen, 2003). The epidemiologic literature has consequently placed great emphasis on the approach to evaluating evidence to determine if a factor can be considered to cause disease. The approaches that have been developed for evaluating causality of associations also draw on multiple lines of scientific evidence; epidemiologic evidence alone is generally not regarded as sufficient for establishing causality (Last, 2000).

Making causal inferences from observational data, in combination with other relevant forms of data, can be a challenging task that requires expert judgment regarding the likely sources and magnitude of confounding, together with judgment about how well the existing constellation of study designs, results, and analyses address this potential threat to inferential validity. This judgment also needs to incorporate a broader assessment of the evidence, evaluating whether a causal effect has support in the existing knowledge of the underlying biologic process. To aid this judgment, criteria for determining a cause have been proposed by many philosophers and scientists over the centuries.

In biomedical research, the first criteria came following the discovery of bacteria during the nineteenth century. A method was then needed for judging if an organism caused a particular disease. The first criteria put forward for making this judgment are generally attributed to Robert Koch and his mentor Jacob Henle, although Koch also acknowledged Eugene Klebs. Evans (1993) provided a full accounting of the elaboration of these criteria, now referred to as the Henle-Koch postulates (Table 1-1). The criteria proved valuable for linking infectious agents to infectious diseases, which often have specific clinical features and unique, specific causal agents (e.g., pulmonary tuberculosis and *Mycobacterium tuberculosis*).

These criteria, however, proved unsuitable for establishing the causes of the epidemics of “chronic disease,” including coronary heart disease, chronic lung disease, and cancer, that became the dominant causes of death spanning the twentieth century, as infectious diseases were controlled. Unlike many infectious diseases, these diseases were often found to be associated with multiple factors, and many cases could not be clearly linked to any risk factors. The limitations of the Henle-Koch postulates were recognized as the results of the first wave of epidemiologic studies on the chronic diseases were reported. In 1959, Yerushalmy and Palmer proposed criteria for evaluating possible etiologic risk factors for chronic diseases that acknowledged the need for evidence of increased risk in exposed persons and for handling the nonspecific causation of these diseases. Lilienfeld (1959) and Sartwell (1960), discussing the article, added the consideration of dose-response, the strength of the association, its consistency, and its biologic plausibility.

The most widely cited criteria in epidemiology and public health more generally were set forth by Sir Austin Bradford Hill in 1965 (Table 1-2) (Hill, 1965). Five of the nine criteria he listed were also put forward in the 1964 Surgeon General’s report (US Department of Health Education and Welfare—DHEW, 1964) as the criteria for causal judgment: consistency, strength, specificity, temporality, and coherence of an observed association. Hill also listed biologic gradient (dose-response), plausibility, experiment (or natural experiment), and analogy. Many of these criteria had been cited in earlier epidemiologic writings (Lilienfeld, 1959; Yerushalmy and Palmer, 1959; Sartwell, 1960); and Susser and others have refined them extensively by exploring their justification, merits, and interpretations (Susser, 1973, 1977; Kaufman and Poole, 2000).

Hill (1965) clearly stated that these criteria were not intended to serve as a checklist.

Here are then nine different viewpoints from all of which we should study association before we cry causation. What I do not believe... is that we can usefully lay down some hard-and-fast rules of evidence that *must* be obeyed before we accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis, and none can be required as a sine qua non. What they can do, with greater or less strength, is to help us to make up our minds on the fundamental

<table>
<thead>
<tr>
<th>Table 1-1. Henle-Koch Postulates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The parasite occurs in every case of the disease in question and under circumstances that can account for the pathologic changes and clinical course of the disease.</td>
</tr>
<tr>
<td>2. It occurs in no other disease as a fortuitous and nonpathogenic parasite.</td>
</tr>
<tr>
<td>3. After being fully isolated from the body and repeatedly grown in pure culture, it can induce the disease anew.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 1-2. Sir Austin Bradford Hill’s Causal Criteria: Aspects of Association to Be Considered Before Deciding on Causation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength</td>
</tr>
</tbody>
</table>

*Source: Hill (1965).*

---

**Figure 1-1. Conceptual scheme for the causes of a hypothetical disease.** *(Source: Rothman, 1976.)*
question—is there any other way of explaining the facts before us, is there any other answer equally, or more, likely than cause and effect? (Hill 1965, p. 299)

All of these criteria were meant to be applied to evidence related to an already established statistical association; if no association has been observed, these criteria are not relevant. Hill explained how if a given criterion were satisfied it strengthened a causal claim. Each of these nine criteria served one of two purposes: as evidence against competing noncausal explanations or as positive support for causal ones. Noncausal explanations for associations include chance; residual or unmeasured confounding; model mis specification; selection bias; errors in measurement of exposure, confounders, or outcome; and issues regarding missing data (which can also include missing studies, such as publication bias). The criteria are discussed below.

Consistency
The criterion consistency refers to the persistent finding of an association between exposure and outcome in multiple studies of adequate power and carried out by different investigators in studies involving different persons, places, circumstances, and times. Consistency can have two implications for causal inference. First, consistent findings make unmeasured confounding an unlikely alternative explanation to causation for an observed association. Such confounding would have to persist across diverse populations, exposure opportunities, and measurement methods. The confounding is still possible if the exposure of interest were strongly and universally tied to an alternative cause, as was claimed in the form of the “constitutional hypothesis” put forward in the early days of the smoking-disease debate (US Department of Health Education and Welfare—DHEW, 1964). This hypothesis held that there was a constitutional (i.e., genetic) factor that made people more likely to both smoke and develop cancer. Thus, consistency serves mainly to exclude the possibility that the association is produced by an ancillary factor that differs across studies but not one factor that is common to all or most of them (Rothman and Greenland, 1998).

The second implication of the consistency criterion is to reduce the possibility of a chance effect by increasing the statistical strength of an association through the accumulation of a large body of data. Consistency does not include the qualitative strength of such studies, which Susser subsumed under his subsidiary concept of “survivability,” relating to the rigor and severity of tests of association (Susser, 1991).

Strength of Association
Strength of association includes two dimensions: the magnitude of the association and its statistical strength. An association strong in both aspects makes the alternative explanations of chance and confounding unlikely. The larger the measured effect, the less likely it is that an unmeasured or poorly controlled confounder could account for it completely. Associations that have a small magnitude or weak statistical strength are more likely to reflect chance, a modest degree of bias, or unmeasured weak confounding. However, the magnitude of association is reflective of underlying biologic processes and should be consistent with understanding the role of the risk factor in these processes. Either a strong or a weak effect might be considered plausible based on knowledge of the underlying processes.

In the example of active smoking and lung cancer, the relative risks listed in the first Surgeon General’s Report (US Department of Health Education and Welfare—DHEW, 1964) were notably elevated in men, reaching as high as 10 or more. At that time, other causes of lung cancer, including air pollution and occupational agents, had been identified. However, for the general population, the risks from these factors were far lower, making them unsatisfactory as potential confounders, leading to the observed association of active smoking with lung cancer.

Passive smoking, by contrast, has a far smaller effect on lung cancer risk. Comparing persons with greater and lesser exposures (e.g., never-smoking women married to smokers compared with never-smoking women married to never-smokers). The 1986 report of the U.S. Surgeon General (US Department of Health and Human Services—USDHHS, 1986) concluded that passive smoking does cause lung cancer. The magnitude of the effect was small in most of the studies, as anticipated on a biologic basis, but within a plausible range. The relative risk associated with marriage to a smoker has been estimated to be 1.2 in a recent meta-analysis (International Agency for Research on Cancer—IARC, 2002).

Specificity
Specificity has been interpreted to mean both a single (or few) effect(s) of one cause or no more than one possible cause for one effect. In addition to specific infectious diseases caused by specific infectious agents, other examples include asbestos exposure and mesothelioma and thalidomide exposure during gestation and the resulting unusual constellation of birth defects. This criterion is rarely used as it was originally proposed, having been derived primarily from the Henle-Koch postulates for infectious causes of disease (Susser, 1991). When specificity exists, it can strengthen a causal claim, but its absence does not weaken it (Sartwell, 1960). For example, most cancers are known to have multifactorial etiologies; many cancer-causing agents can cause several types of cancer, and these agents can also have noncancerous effects.

When considering specificity in relation to the smoking—lung cancer association, the 1964 Surgeon General’s report (US Department of Health Education and Welfare—DHEW, 1964) provides a rich discussion of this criterion. The committee recognized the linkage between this criterion and strength of association and offered a symmetrical formulation of specificity in the relation between exposure and disease; that is, a particular exposure always results in a particular disease, and the disease always results from the exposure. The committee acknowledged that smoking does not always result in lung cancer and that lung cancer has other causes. The report noted the extremely high relative risk for lung cancer in smokers and the high attributable risk, and it concluded that the association between smoking and lung cancer has “a high degree of specificity.”

Temporality
Temporality refers to the occurrence of a cause before its purported effect. Temporality is the sine qua non of causality, as a cause clearly cannot occur after its purported effect. Rothman (1986) emphasized that temporality is the only one of the criteria that must be fulfilled for an association to be considered causal. Any question about a temporal sequence seriously weakens a causal claim; but establishing temporal precedence is by itself not strong evidence in favor of causality.

Coherence, Plausibility, and Analogy
Although the original definitions of coherence, plausibility, and analogy were subtly different, in practice they have been treated essentially as one idea: that a proposed causal relation should not violate known scientific principles, and that it be consistent with experimentally demonstrated biologic mechanisms and other relevant data, such as ecologic patterns of disease (Rothman and Greenland, 1998). In addition, if biologic understanding can be used to set aside explanations other than a causal association, it offers further support for causality. Together, these criteria can serve to both support a causal claim (by supporting the proposed mechanism) and refute it (by showing that the proposed mechanism is unlikely).

Biologic understanding, of course, is always evolving as scientific advances make possible ever deeper exploration of disease pathogenesis. For example, in 1964 the Surgeon General’s committee found the causal association of smoking with lung cancer to be biologically plausible based on knowledge of the presence of carcinogens in tobacco smoke and animal experiments. Nearly 40 years later, this association remains biologically plausible, but that determination rests not only on the earlier evidence but on more recent findings that address the
Biologic Gradient (Dose-Response)

The finding of a graded increase in effect with an increase in the strength of the possible cause provides strong positive support in favor of a causal hypothesis. This is not just because such an observation is predicted by many cause-and-effect models and biologic processes but, more importantly, because it makes most noncausal explanations highly unlikely. If some factor other than that of interest explains the observed gradient, the unmeasured factor must change in the same manner as the exposure of interest. Except for confounders that are closely related to a causal factor, it is extremely difficult for such a pattern to be created by virtually any of the noncausal explanations for an association listed earlier. The finding of a dose-response relation has long been a mainstay of causal arguments in smoking investigations; virtually all health outcomes causally linked to smoking have shown an increase in risk and/or severity with an increase in the lifetime smoking history. This criterion is not based on any specific shape of the dose-response relation.

Experiment

The criterion “experiment” refers to situations where natural conditions might plausibly be thought to imitate conditions of a randomized experiment, producing a “natural experiment” whose results might have the force of a true experiment. An experiment is typically a situation in which a scientist controls who is exposed in a way that does not depend on any of the subject’s characteristics. Sometimes nature produces similar exposure patterns. The reduced risk after smoking cessation serves as one such situation that approximates an experiment; an alternative noncausal explanation might posit that an unmeasured causal factor of that health outcome was more frequent among those who did not stop smoking than among those who did. The causal interpretation is further strengthened if risk continues to decline in former smokers with increasing time since quitting. Similar to the dose-response criteria, observations of risk reduction after quitting smoking have the dual effects of making most noncausal explanations unlikely and supporting the biologic model that underlies the causal claim.

APPLYING THE CAUSAL CRITERIA

The greater the extent to which an association fulfills the previous criteria, the more difficult it is to offer a more compelling alternative explanation. Which of these criteria may be more important and whether some can be unfilled and still justify the causal claim is a matter of judgment. Temporality, however, cannot be violated. When there is a still incompletely understood pathogenic mechanism, the causal claim might still be justified by strong direct empirical evidence of higher lung cancer rates in smokers (i.e., strong, consistent associations). Moderate associations (e.g., relative risk of 1–2) in only a few studies, without adequate understanding of potential confounders or with weak designs, might result in a suspicion of causal linkage.

The process of applying the criteria extends beyond simply lining up the evidence against each criterion, although there is evidence that epidemiologists tend to use the evidence in neither a consistent nor comprehensive manner (Weed and Gorelic, 1996). Rather, the criteria should be used to integrate multiple lines of evidence coming from chemical and toxicologic characterizations of tobacco smoke and its components, epidemiologic approaches, and clinical investigations. Those applying the criteria weigh the totality of the evidence in a decision-making process that synthesizes and, of necessity, involves a multidisciplinary judgment.

The 1986 Report of the Surgeon General (US Department of Health and Human Services—USDHHS, 1986) concluded that passive smoking causes lung cancer, a conclusion that has proved momentous in its implications. This report also based its evaluation of the evidence on the causal criteria. A clear distinction was made between the evidence on active smoking and that expected from the much lower carcinogen doses arising from passive smoking. Biological plausibility was emphasized, including the substantial evidence on lung cancer risk in active smokers. This causal conclusion has been reaffirmed in all subsequent reports (Samet and Wang, 2000; International Agency for Research on Cancer—IARC, 2002).

EMERGING ISSUES IN CAUSAL INFERENCE AND CANCER

Perhaps the most challenging and exciting issue facing cancer scientists now and in the future is the prospect of understanding the processes of cancer development at the molecular level. However, with a richer understanding of basic mechanisms comes concomitant complexity in the concept and determination of cause. Biomarkers can serve as indicators of exposure, dose, susceptibility, or effect, each of which can elevate the cancer risk, albeit via quite different routes (Links et al., 1995). Similarly, the mechanisms by which various genes affect cancer risk are diverse, from genes that directly modulate tumor growth to others responsible for cellular homeostasis, DNA repair, genetic stability, or a host of interrelated functions that protect the cell against damage from somatic or environmental factors or affect its repair capacity when damage occurs (Vineis and Porta, 1996; Hussain and Harris, 1998).

It is interesting to consider the implications of this kind of knowledge for causal inference in cancer. The most obvious change is that we now understand the biologic basis of action of long-established carcinogens, such as smoking, chemotherapy, and various chemical agents. Mechanistic explanations of how environmental exposures have a carcinogenic effect provide the basis for increased confidence that any given association between the exposure and cancer incidence is justifiably labeled causal. Molecular “signatures” of specific exposures (e.g., p53 CpG hotspots) (Greenblatt et al., 1994) are defining the relevant effects of certain exposures more precisely and are also making increasingly possible what could not be done before: establish causal connections between exposure and disease on the individual level.

Second, by understanding better what mediates risk due to exposures, we are increasingly able to identify subpopulations of individuals at substantially different degrees of risk from an exposure (Shields...
and Harris, 2000). This risk heterogeneity can be due to genetic or somatic variations that affect the absorption, metabolism, or cellular effect of an agent or the host's susceptibility to those effects (Links et al., 1995). The smaller size of these subpopulations, however, poses greater challenges to epidemiologic approaches to risk assessment: the finer the risk stratification based on mechanistic criteria, the more difficult it is to confirm it with epidemiologic methods (Slattery, 2002; Moore et al., 2003).

Third, with the opening of the cancer “black box” comes greater recognition of the extraordinary complexity of the carcinogenic process at the molecular level. This is seen not only with external agents dependent on genes for their effects but with environmental modification of gene expression (e.g., through DNA methylation) (Moore et al., 2003), gene functions interacting in myriad ways, and perhaps most strikingly genetic determinants of exposure, as shown with developing understanding of the biologic bases of nicotine, alcohol, and drug addiction (Shields et al., 1993; Greenblatt et al., 1994; Hemby, 1999). This complexity poses significant problems for population-based risk assessments and causal inference because it raises, more acutely than with conventional risk factors, questions about which genes lie in the causal pathway (those that do not should not be considered covariates), which genes or biomarkers are necessary for the effects or function of others (i.e., there may be many biologically based high level interaction terms), whether an observed association represents a direct or indirect causal effect, and whether it is meaningful to talk about binary exposure-outcome or gene-outcome relations when the components of these complex networks either cannot be changed individually or changed at all (Taioli and Garte, 2002).

Fourth, although the functions of many genes and gene products are understood, many more are not. The current ability to perform mass screening of potential causal factors via high throughput, genomic and proteomic technologies is far outstripping our ability to explain what we find; nearly all the problems related to multiplicity (“data dredging”) arise in many of these investigations (Reiner et al., 2003). Epidemiologists have long faced this problem but never on this scale, with literally hundreds or even thousands of potential markers measured and analyzed, in all combinations, sometimes in only tens or hundreds of subjects. The weakness of findings subject to these problems is not often fully appreciated, and researchers are making claims about possibilities of screening, treatment, or prevention before findings are replicated in independent data sets.

A final issue that is often neglected in this rush toward exploring new potential causes is that of measurement (Little et al., 2002). Many of the techniques and assays used to identify metabolic products, genes, and gene products are relatively new and not standardized. Understanding the reliability and validity of these techniques is an arduous process that often does not get sufficient attention, yet it is critical for distinguishing likely spurious claims from those that are well grounded.

As we come closer to understanding cancer on the molecular level, many have raised the possibility of individual risk prediction (Vines, 1997; Hussain et al., 2001). Some commentators have suggested that mechanistic understanding may ultimately render population studies irrelevant. However, it is instructive to consider the case of infectious diseases. Understanding the basic mechanisms of infectious disease has not eliminated the need to study disease patterns on a population level, and the same will likely be true for cancer (Nevins et al., 2003). There are many reasons to believe that population-based studies will be as important in the future as they are now, albeit perhaps focused on different kinds of questions. To understand a mechanism after a cancer occurs is not to predict it; early steps in the process, when the disease can be prevented, will necessarily be less than 100% predictive; and defining optimal groups for screening and early intervention will still require population-based data. Risk groups must be defined using far fewer factors than we know are operating at the molecular level, the latter being almost unique for an individual.

The molecular revolution in cancer may ultimately force a merging of two “schools” of causal inference: the probabilistic, chronic disease model that the Hill criteria addressed and the more mechanistic, determinist models used for infectious disease, for which the Henle-Koch criteria were devised (Fredericks and Relman, 1996). Nowhere is this seen more clearly than with viral carcinogenesis. The Henle-Koch criteria were based on the nineteenth century understanding of bacterial disease causation and are poorly suited for viral disease mechanisms or even for infectious disease outcomes. Efforts to refashion the Henle-Koch criteria for the new era of molecular medicine (Fredericks and Relman, 1996; Vines and Porta, 1996) have shown that it is extraordinarily difficult to outline a set of experimental conditions that all known pathogens—not to mention new pathogens with different mechanisms—must satisfy to justify causal claims for new infectious diseases.

In the case of viral carcinogenesis, the situation is even more complex because the final disease is not a direct manifestation of an infectious process. Numerous viral agents have been linked to cancer with varying degrees of certainty: Epstein-Barr virus and Burkitt’s lymphoma (Pagano, 1999), human papillomavirus and cervical cancer (Bosch and de Sanjose, 2003), SV40 and mesothelioma (Carbone et al., 1997; Klein et al., 2002). The evidential basis of these claims includes findings that might implicate the virus in individual cancer cases (e.g., finding viral genetic sequences or viral-specific proteins in tumor tissue) and traditional epidemiologic evidence (e.g., high incidence in persons with evidence of viral exposure). However, as both these and other examples have shown, there is no single molecular finding that definitively implicates a virus as a cause of cancer. The variety and complexity of mechanisms by which viruses can directly (by inducing oncogenic changes) or indirectly (by increasing host susceptibility to exogenous agents) raise cancer risk seems to defy a set of causal criteria designed specifically for viral agents or that are based on any specific mechanism. Therefore, from the standpoint of causal inference, it may be best to consider viral agents under the same umbrella as toxic exposures and other environmental causes of cancer. The fact that the pathways from viral infection to cancer appearance often share components (e.g., p53 inactivation) with those of noninfectious carcinogenic exposures supports this view.

It seems unlikely that the need for population-based risk estimates and causal inference will disappear from cancer research, just as it has not in infectious disease. However, we are entering an era where the relative strength of the “twin pillars” of causal inference—knowledge derived from empirical, population-based patterns and that based on understanding of biologic mechanisms in individuals—will tilt farther toward the mechanistic end, requiring less proof from populations and more from the laboratory. One challenge for causal inference in the future will be how best to integrate these various forms of evidence and how to assemble groups with the sufficient interdisciplinary expertise to assess them.

References