Chapter 9
Cause of Disease and Causal Inference

Li Ye

Key Points

- In epidemiology, cause and causal inference are used to explore the etiology of risk factors for diseases at a population level.
- A causal model is a concise and conceptual graphic that describes the relationship between cause and disease.
- Most epidemiologic study designs can be used for evaluating causation. The strength of these designs to evaluate causation varies.
- Mill's canons represent logical strategies for inferring a causal relationship.
- Hill's criteria are a list of guidelines to distinguish causal and noncausal associations; these criteria have been widely used and are the best known criteria for assessing causal inference.

9.1 Introduction

One of the major focuses of epidemiology is to find the causes of diseases or events. Understanding the causes of diseases is important not only for correct diagnoses and treatments but also for effective prevention and control strategies. Therefore, cause of disease and causal inference—the process by which we identify the cause of disease are essential in both clinical and preventive medicine. In epidemiology, cause and causal inference are used to explore the etiology of or risk factors for diseases as well as their impact on the development of disease at the population level, which can provide unique insights into the etiology of the disease and lead to a population-level understanding of the disease. This chapter describes the epidemiologic concept of cause and the approaches to causal inference.

L. Ye (✉)
School of Public Health, Guangxi Medical University, Nanning, China

C. Wang, F. Liu (eds.), *Textbook of Clinical Epidemiology*,
https://doi.org/10.1007/978-981-99-3622-9_9

9.2 Cause of Disease in Epidemiology

9.2.1 The Concept of Cause in Epidemiology and its Development History

There are many definitions of cause in epidemiology. The following widely accepted definition is from Abraham Lilienfeld: a causal relationship would be recognized to exist whenever evidence indicates that the factors from part of the complex of circumstances that increase the probability of the occurrence of disease and that a diminution of one or more of these factors decrease the frequency of that disease. Another definition from Kenneth Rothman and Greenland [10] is also widely accepted due to its simplicity and clarity: an event, condition or characteristic, or a combination of these factors that play an essential role in producing an occurrence of the disease. Cause is an important concept in epidemiology. There are many other synonyms to describe cause, including causal agency, determinant, risk factor, exposure, etiological factor, etiological agent, etc. In epidemiology, cause is often referred to as a risk factor, which means the factor that increases the risk of disease.

The cause of disease has long been explored. The most ancient idealism attributed the occurrence of diseases to the god or devil. In the fourth century BC, Hippocrates, the father of medicine, considered that diseases occurred because of the imbalance of "four body humors." In the fifth century, Chinese ancestors founded a materialistic view of the cause, and they proposed diseases were from the imbalance of "Yin-Yang" or "Five elements (wood, earth, water, metal, fire)."

In the later nineteenth century, at the height of the era of germ theory, Robert Koch, the founder of modern bacteriology, proposed Koch's postulates, which include four generalized principles for determining whether a specific microorganism causes a specific disease. Koch's postulates contributed greatly to the formation of the concept of cause in epidemiology because identifying the microorganism was equivalent to identifying the cause of the disease. In fact, the discipline of epidemiology as well as the concept of the cause of disease originated from etiology and epidemic studies on communicable diseases, among which the germ theory and Koch's postulates represent landmark achievements.

However, Koch's postulates cannot explain the causes for most diseases, especially noncommunicable diseases, which have replaced communicable diseases as the main threat to human health since the middle of the twentieth century. More recently, the epidemiologic studies have focused more on the probability and multicausality of the occurrence of diseases, which finally led to the formation of a modern concept of cause, as described at the beginning of this subsection.

9.2.2 *Classification of Cause*

In modern epidemiology, the concept of cause actually means "multicausality". Most diseases, whether communicable or noncommunicable, have more than one cause. Since the definition of cause, either by Lilienfeld or by Rothman, means that any "factor" or any "event, condition or characteristic" plays a role in affecting the occurrence of the disease, the cause in epidemiology covers a wide range of factors, including individual genetics, physiological influences, environmental influences, social structure, etc. According to the source of these factors, we can divide the causes into two general categories: host factors and environmental factors (Table 9.1). Host factors refer to various characteristics that are related to people or an individual, such as genetics, immune status, age, sex, race, and behavior. (Table 9.1). Environmental factors mainly include biological, physical, chemical, and social factors (Table 9.1).

Table 9.1 Classification of the cause of disease

Factor (cause)	Description
Host factors	
1. Genetics	Chromosomal disorder, single gene disorder, polygenetic disorders, etc.
2. Immune status	It involves in the occurrence of most diseases, both communicable and noncommunicable
3. Age and sex	People of different age or sex may be susceptible to different diseases
4. Race	Occurrence of disease has difference in race
5. Personality	Temperament, psychological status, psychiatric status, etc. may have effects on the occurrence or progression of diseases
6. Behavior	Bad behaviors or habits such as smoking, drinking, poor diet, lack of exercise, unsafety sexual behaviors, drug abuse, noncompliance with traffic laws, etc.
Environmental factors	
1. Biological	Pathogenic microorganisms (bacteria, viruses, rickettsiae, mycoplasmas, chlamydiae, spirochetes, actinomyces, etc.); parasites (protozoa, worms, insects, etc.); venomous animals and poisonous plants (snakes, ergot, mushrooms, etc.)
2. Physical	Temperature, humidity, altitude, noise, light, vibration, radiation, dust, fire, etc.
3. Chemical	Pollution, agricultural chemicals, food additives, microelement, heavy metal, etc.
4. Social	Social system, socioeconomic level, war, disaster, education, religion, living condition, lifestyle, occupation, family relationship, etc.

9.2.3 Causation Models

A causal model is a concise and conceptual graphics that describes the relationship between cause and disease. During the development of etiology, different causal models were proposed based on contemporary understanding of the diseases in different historical periods, which made a great contribution to the formation of the modern concept of cause. Casual models can be used to illuminate the association between cause and disease as well as the relationship between multiple causes of a disease and to provide direction or clues to find a new cause. Essentially, the aim of causal models is to find causes and elucidate the dominant cause and ultimately determine the best prevention or intervention strategy. The most representative causal models are the triangle model, the wheel model, the chain of causation model, and the web of causation model.

9.2.3.1 Triangle Model

In 1954, John Gordon summarized the knowledge about the epidemiologic etiology of diseases at that time and put forward an epidemiologic triangle model (epidemiologic triad) to describe the relationships between multifactorial causes and a disease, especially communicable disease. The model considers that host factors (age, sex, race, genetic profile, immune status, etc.), agents (biologic pathogens, chemical, physical, nutritional agents, etc.), and environmental factors (temperature, humidity, crowding, housing, water, food, radiation, pollution, noise, etc.) are the troika of a disease. These three aspects are indispensable for the occurrence of a disease and have an equal role in the occurrence of disease. Hence, the relationships can be described as an equilateral triangle (Fig. 9.1). The three kinds of factors interact and restrict each other, and thus, a dynamic balance exists that makes the occurrence of disease in a stable state. Once the balance is disturbed, the occurrence of disease increases or decreases. The triangle model is helpful even today for finding the cause of communicable diseases and controlling the epidemic. However, it is basically unsuitable for the description of noncommunicable diseases.

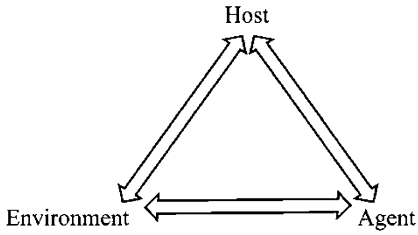


Fig. 9.1 Epidemiologic triangle model

9.2.3.2 Wheel Model

In the middle of the twentieth century, noncommunicable diseases became the main threat to humankind. However, there is no obvious or absolute agent for most noncommunicable diseases, as a pathogen agent for communicable diseases. It is very difficult to describe the relationship between cause and noncommunicable disease using a triangle model. In 1985, Mansner and Kramer proposed the wheel model based on the triangle model. In this model, host factors play the core role in the occurrence of disease and are located in the center of the wheel, with genetic factors as the core of the center (Fig. 9.2). The host center is surrounded by three kinds of environmental factors, including biological, social, and physical environmental factors. The main difference between the wheel model and the triangle model is that the wheel model considers that different factors have different importance for the occurrence of disease. Therefore, the area sizes of the center (host factors) and surrounding parts in the wheel (biological, social, and physical factors, respectively) can be adjusted to reflect the importance of different factors. The wheel model emphasizes the core role of host factors as well as the influencing effects of environmental factors. It is considered to be better than the triangle model and suitable for both communicable and noncommunicable diseases. However, the wheel model came from etiology knowledge in the 1980s and could not truly reflect the complex interactions between various factors. It is still limited for many noncommunicable diseases, especially chronic diseases.

9.2.3.3 Chain of Causation Model

In multicausality theory, there are multiple causes of communicable and noncommunicable diseases. The multiple causes or risk factors can always be displayed in the form of a chain. Some factors are direct or proximal or most immediate causes, and others are indirect or distal causes. Some factors are

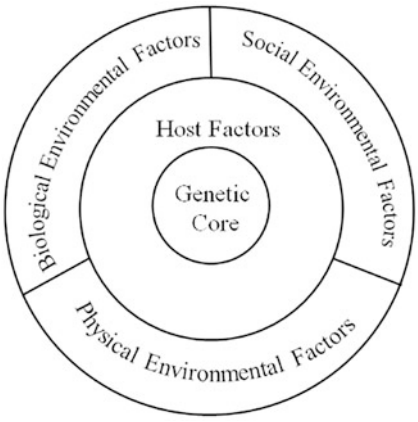


Fig. 9.2 Causation wheel model

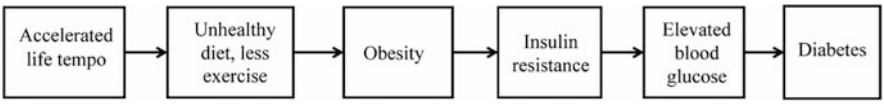


Fig. 9.3 Chain of causation (diabetes)

independent causes; however, others are dependent causes that interact with other causes. To interpret the association of different causes and final disease as well as the complex relationship among multiple causes, a model of "chain of causation" was proposed to describe the causes in the form of a chain. For example, an accelerated life tempo can lead to an unhealthy diet or less exercise and then obesity, followed by insulin resistance, which often results from obesity and further results in elevated blood glucose. Finally, diabetes occurs when blood levels of glucose become chronically elevated (Fig. 9.3). It is worth mentioning that removing any factor in the chain can block the whole chain and thus prevent the occurrence of disease (Fig. 9.3).

9.2.3.4 Web of Causation Model

In some cases, a disease has several and interrelated chains of causation. The causation chains of the disease link and interplay with each other, and thus constitute a complex network. MacMahon proposed the web of causation model to describe the complex relationships between causes and disease as well as the interlacing chains of causation.

For example, the causation of liver cancer can be described as a network or a web. The four chains of causes of liver cancer consist of biological factors, physical and chemical factors, behavioral factors, and genetic factors (hereditary susceptibility). Multiple factors of the four chains also interact with each other and form a network, thus ultimately leading to the occurrence of liver cancer (Fig. 9.4).

9.2.4 Sufficient Cause and Necessary Cause

Modern epidemiology considers that the relationships between cause and effect are multiple and complex. A given disease can be caused by many factors; however, a single factor is not enough to cause the disease, as joint action from other causes is necessary. Obviously, the role and importance of different factors are different in the occurrence of a disease. From the logical view of cause and effect, all effects have sufficient and necessary conditions; thus, cause in epidemiology can also be divided into sufficient and necessary cause.

In 1976, Kenneth Rothman used a sufficient-component causal model (Fig. 9.5) to explain the complex relationship between cause and effect. Rothman proposed that a sufficient cause is a factor or a combination of several factors that will

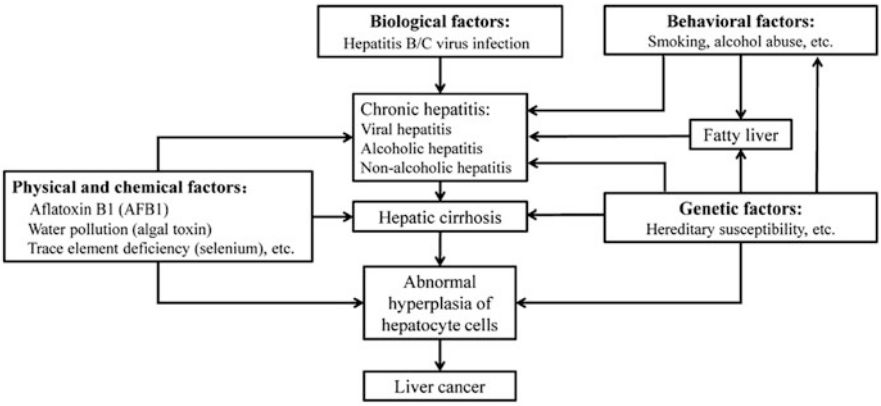
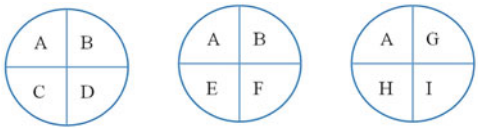


Fig. 9.4 Web of causation (liver cancer)

Fig. 9.5 Sufficient cause and necessary cause



Sufficient cause I Sufficient cause II Sufficient cause III

inevitably cause disease. A component cause is a factor that contributes to the occurrence of disease but is not sufficient to cause disease on its own. A necessary cause is any agent that is required for the occurrence of disease (for example, cholera bacillus for cholera occurring); without necessary cause, the disease will not occur. For instance, the three sufficient causes (I, II, III) shown in Fig. 9.5, comprise 4 component causes. In this figure, there are three sufficient causes (I, II, III), and A, B, C, D, E, F, G, H, and I are component factors. Because A is present in all three sufficient causes, it is a necessary cause.

The sufficient-component causal model interprets two paradoxes of the causation theory in epidemiology. First, why does a given disease occur without a specific cause? For example, alcohol abuse is the cause of cirrhosis; however, individuals who never drink may also develop cirrhosis. The possible reason is that cirrhosis develops through other sufficient causes, such as hepatitis B virus infection (Fig. 9.4). Second, why does a disease fail to occur in the presence of a specific cause? For example, smoking causes lung cancer; however, many smokers never develop lung cancer in their lifetime. This phenomenon can be explained by the fact that smoking is not a sufficient cause of lung cancer. In reality, most identified causes for noncommunicable chronic diseases are neither necessary nor sufficient. For example, hypertension is neither a necessary cause nor a sufficient cause for cardiovascular disease. Nevertheless, every component cause is necessary for sufficient cause that contains it. Removal of any component cause is equal to the removal of

the sufficient cause that contains this component cause, which is an important strategy for the prevention of a disease.

9.3 Epidemiologic Methods of Causation

9.3.1 Epidemiologic Study Designs for Causation

Etiological studies in epidemiology usually contain several common steps. The first step is to find the influencing factors that are associated with a disease or to develop a cause-and-effect hypothesis, usually by descriptive or analytical epidemiologic studies; the second step is to test the hypothesis often using analytical studies; and the last step is to verify the hypothesis, usually by experimental epidemiologic studies.

Most epidemiologic study designs can be used for evaluating or establishing causation. However, the strength of these designs to evaluate causation is different. Table 9.2 outlines the relative strength of the different study designs in establishing causation. These study designs have been introduced in prior chapters, and their use in providing evidence for causation will be described as follows.

9.3.1.1 Descriptive Studies

The start of causation is to develop a cause-and-effect hypothesis. Descriptive studies are always used to generate hypotheses. Descriptive studies mainly include case reports, case series, cross-sectional studies, and ecological studies. Case reports and case series are useful for developing a hypothesis based on analysis of the characteristics of patients or case groups. Cross-sectional studies are always used to describe the distribution of disease in different populations, and the pattern or trend of disease occurrence over time or by geographic area, which can provide the clues regarding influencing factors. Ecologic studies explore the association of influencing factors and disease at the population or region level, which can also provide clues for influencing factors that cannot be measured at the individual level, for example, air

Table 9.2 The strength of evidence for causation by different epidemiologic study designs

Type of study design	Strength of evidence in causation
Randomized controlled trials	Strong
Nonrandomized controlled studies	Moderate
Cohort studies	Moderate
Case-control studies	Moderate
Cross-sectional studies	Weak
Ecologic studies	Weak
Case reports	Weak

pollution. Generally, the strength of descriptive studies to evaluate causation is weak compared with analytical studies or experimental studies due to a lack of evidence on the time sequence of events. Of all descriptive studies, the weakest for causation is case reports because they have neither defined populations nor comparison groups. Nevertheless, when causal relationships have already been established, well-designed descriptive studies, especially cross-sectional studies with multiple time points or time series studies, can be very useful to quantify the effects of cause.

9.3.1.2 Analytical Studies (Case-Control Studies, Cohort Studies)

Analytical studies, mainly including case-control studies and cohort studies, are more reliable methods to form a hypothesis than descriptive studies. Analytical studies can also be used for hypothesis testing. Case-control studies, which are mainly used to confirm the association between factors and disease, compare the exposure levels between the case group and the control group. Because the research direction of case-control studies is from effects (diseases) to causes (factors), this study design is vulnerable to various biases. Cohort studies are either prospective or retrospective, and they can test hypotheses more effectively by comparing the incidence rates of exposure groups with control groups, and directly calculating the relative risk (RR) of factors in the temporal order of cause and effect. Well-conducted cohort studies are a better design for causation than case-control studies because the former can minimize various biases, including selection, information, and confounding biases.

9.3.1.3 Experimental Studies

Experimental studies include clinical trials, field trials, and community trials. Clinical trials are most frequently conducted among patients, with the aim of evaluating the efficacy of a new treatment or medicine. Therefore, clinical trials are known as a robust and reliable method to test or verify hypotheses, especially clinical randomized controlled trials (RCTs), which are considered the gold standard to evaluate a new treatment or medicine and the most rigorous method for hypothesis testing. Nevertheless, RCTs are subjected to many constraints, such as ethical issues, strict inclusion criteria, parallel control, and strict randomization, which greatly limit the feasibility of RCTs in causation studies. Quasi-experiments that lack parallel control or randomized assignment are also used in epidemiologic etiology, with less strength to evaluate causation. Other experimental studies, including field and community trials, are seldom used to study causation. Field trials mainly involve people who are disease-free, with the aim of preventing the occurrence of diseases. Community trials are conducted at the level of the community instead of the individual level. Therefore, although experimental studies have strong strength to test and verify hypotheses, most of the causative evidence so far has not come from this study design but comes from observational studies such as descriptive and analytical studies. For

example, most of evidence about the effects of smoking on health comes from case-control and cohort studies.

9.3.2 Mill's Canons-the Logical Basis of Causation

The causa models mentioned in Sect. 9.2.3 of this chapter are mainly used to describe the relationships between the factors and diseases or between different risk factors. They cannot be used as a method to find causes or test causation. Mill's canons were proposed by philosopher John Stuart Mill in 1843, intended to illuminate a causal relationship between a circumstance and a phenomenon, which provides the logical basis of causation studies. In epidemiology, Mill's canons provide certain guidance for causation, especially the development of hypotheses. The canons with minor adjustments constitute five methods for the induction of hypotheses.

9.3.2.1 Method of Agreement

This method means that factors in common among different instances of a disease are perhaps the cause or a necessary part of the cause of the given disease. In other words, if two or more instances of a disease under investigation have only one factor in common, which is likely to be the cause of the given disease. For example, one school had an outbreak of diarrhea. It was found that all the students with diarrhea had consumed soy milk in the same canteen in the morning, so soy milk may be the cause of diarrhea.

However, in actual conditions, it is difficult to obtain only "one common factor". There may be a few other factors shared by patients with the same disease, but most of the factors are not the cause of the disease. In the example mentioned above, most students with diarrhea may have also eaten another food in common in the same cafeteria that morning. Therefore, the method of agreement in this example is actually not sure that soy milk may be the cause of outbreaks of diarrhea. Generally, a hypothesis cannot be formed by one method.

9.3.2.2 Method of Difference

If some instances in which the disease occurs, and other instances in which the disease does not occur, they have all other factors in common except one existing only in the former. That one may be the cause or a necessary part of the cause of the disease. The method of agreement concerns whether patients share certain common factors. The method of difference compares the differences in certain characteristics between patients and nonpatients. For example, if one student had diarrhea while the

other did not, the only different food that two students consumed in the canteen was soy milk, and the soy milk may be the cause of diarrhea.

Similar to the situation of the method of agreement, in the method of difference, the assumption that "all other factors are the same" between patients and nonpatients is hard to make in practice. In the above example, the two students may be different from each other in many other aspects. The method of difference cannot exclude other factors and hypothesize that only soy milk is the cause of diarrhea.

9.3.2.3 Joint Methods of Agreement and Difference

This method is actually a combination of the method of agreement and the method of difference, however, it is not a simple combination but alternately contains multiround use of two methods. Briefly, if two or more instances in which the disease occurs have only one factor in common, while two or more instances in which the disease does not occur have nothing in common except the absence of the factor commonly existing in the former instances, then, the factor may be the cause, or a necessary part of the cause, of the disease. Generally, the joint methods of agreement and difference are much more likely to find a risk factor than the method of agreement or difference alone. The main reason is that joint methods of agreement and difference essentially introduce contrast in the investigation, which greatly increases the logicality.

We return to the example of soy milk and diarrhea. If all of the students with diarrhea had consumed soy milk, the students without diarrhea must have not consumed soy milk in the same canteen. This is the joint method to indicate that soy milk was likely to be the cause of diarrhea.

9.3.2.4 Method of Concomitant Variations

According to the method of concomitant variations, whatever one event varies in any manner whenever another event varies in some particular manner. The former event is either a cause or an effect of the latter; in other words, these two events are connected through cause-and-effect association. In essence, method of concomitant variations emphasizes dose-dependent relationships. When there is a dose-dependent relationship between two events, cause-and-effect associations are more likely exist.

In the above examples, if students who consumed more soy milk had more severe diarrhea, there was a dose-dependent relationship, namely, concomitant variations, so the probability of the soy milk as the cause of diarrhea was higher.

9.3.2.5 Method of Residue

Suppose a disease is caused by many factors; when you remove the previously known factors as well as the instances of disease caused by those factors, the residue of factors may be the cause for the remaining instances of disease. For example, in 1972, a number of dermatitis cases occurred in Shanghai, China. Possible factors, including industrial waste gas, plant pollen, blood-sucking arthropods, and poisonous moths, as well as dermatitis cases caused by these factors were excluded. The residual factor, Euproctis similis, emerged. Therefore, researchers suspected that this outbreak of dermatitis was caused by Euproctis similis, and finally confirmed this hypothesis.

Although Mill's canons are considered as a logical basis for causation studies and still have certain guidance significance for current epidemiologic etiologies, they have great limitations in actual practice in causation studies. Generally, Mill's canons are suitable to find both sufficient and necessary causes, such as acute infectious agents and to judge strong causal associations. However, for most diseases, especially noncommunicable chronic diseases, the risk factors are almost all nonsufficient and nonnecessary causes, and one disease always has multiple sufficient causes. In these cases, the Mill's canons are not suitable for effectively assessing the causal association. At most, the canons just play a role in the formation of a hypothesis.

9.4 Causal Inference

Causal inference is the term used for the process by which we identify the cause of disease. In other words, it is used to determine whether the observed association is causal. In essence, the relationship between cause and disease is a kind of cause-and-effect association in philosophy. Various risks or exposure factors are the cause, and diseases are the effect. The term association is another important concept in epidemiology.

9.4.1 Association Vs. Causation

Two events, the suspected cause and the effect, obviously must be associated if they are to be determined as causally related. However, not all associations are causal, namely, cause-and-effect associations. Various other associations, including chance association, spurious association, and noncausal association, which are caused by various reasons such as random error, bias, or confounding, should be excluded before a causal association is assessed. Figure 9.6 outlines various associations caused by different reasons.

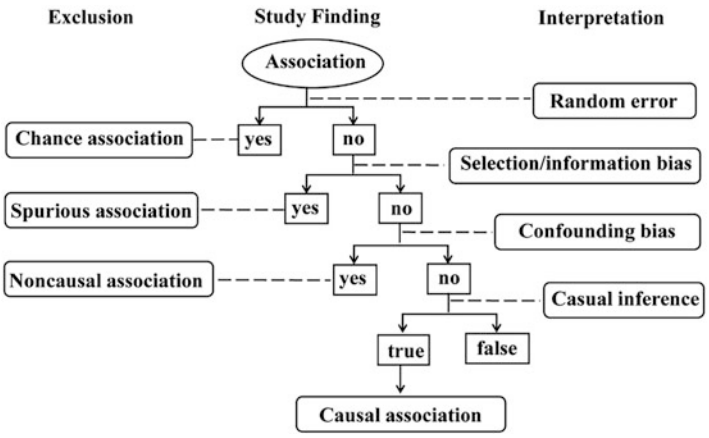


Fig. 9.6 Association and causation

9.4.1.1 Chance Association

First, we need to judge whether an association between two events, for example, an exposure and an outcome, is a statistically significant association rather than a chance association due to random error, e.g., sampling error and random measurement error (Fig. 9.6). In epidemiology, the strength of association can be expressed as rate ratios, odds ratios, or attributable risks, which are introduced in prior chapters. The exclusion of chance association is based on statistical comparison of these indicators between the exposed group (or case group) and the control group. When the *P* value was less than 0.05, we considered that there was a statistical significant difference, namely, a statistically significant association. The *A* value less than 0.05 is defined as a low-probability event, which means that the result is reliable in at over 95% probability, and there is less than a 5% chance that the result is caused by random error. Because random error cannot be avoided and exists in all study designs as well as each step of a study, well-designed and well-conducted studies are essentially important to effectively reduce chance association caused by random error, despite almost no study being perfect in either design or conduct practice.

9.4.1.2 Spurious Association

If the association is statistically significant and the probability of random error is very limited, then we could evaluate whether the association is spurious, which generally comes from nonrandom systematic errors known as selection or information bias (Fig. 9.6). Spurious association means that the association truly exists, but is not true due to selection or information bias. Thus, it is also called as false association. Selection bias is generally caused by the difference in exposure- or outcome-related characteristics between subjects selected for study and those not

selected or between the exposed (or case) group and the control group. For example, in a case-control study on birth defects, the case group of newborns with birth defects, while the control group contained consists of those without birth defects. The collection of exposure information was primarily based on the memory of mothers of the newborns. In information collection, the mothers of newborns with birth defects, who were stimulated by adverse pregnancy were able to recall various exposures during pregnancy in detail, such as taking over-the-counter drugs, fever, or cold. However, mothers in the control group were less likely to make an effort to recall the details and did not respond carefully to the relevant exposure events, because no adverse pregnancy occurred. Therefore, the results obtained may be influenced by recall bias. The association between potential exposure factors and neonatal birth defects may be overestimated, and it may be a false association.

9.4.1.3 Noncausal Association

Even though a true association exists, it is still necessary to know whether the association occurs indirectly by another extraneous factor (called a confounding factor), which always leads to a noncausal association. If confounding was not found, a noncausal association could be excluded and a causal association may exist (Fig. 9.6). Confounding bias is caused by an extraneous factor (confounding factor) that is closely related to both exposure and outcome but not an intermediate link in the causal chain of the exposure and the outcome. Confounding bias can lead to underestimation or overestimation of the association between the exposure and the outcome. For example, investigation may find an association between smoking and alcoholic liver, obviously, which is not reasonable in biological plausibility. The link between smoking and alcoholic liver is a noncausal association because alcohol drinking consumption is a confounding factor, which that is often associated with smoking and directly related to alcoholic liver. Confounding occurs commonly in epidemiologic studies. However, it can be well controlled through careful designs such as matching, restriction, and randomization or through analyses such as standardization, stratification, and multivariate analysis.

9.4.1.4 Causal Association

After excluding chance association, spurious association, and noncausal association caused by random error, selection/information bias, and confounding bias, the association between the exposure and the outcome is likely to be causal, and still needs to be further assessed by various judgments. We call this process causal inference (Fig. 9.6). Among various judgments, the best known and widely used is Hill's criteria. The details of Hill's criteria are introduced in the next subsection.

9.4.2 *Evaluating Causal Association—Hill's Criteria*

The existing association between two events or two variables after exclusion of chance, spurious, and noncausal associations is just probably to be a cause-and-effect association, which is required for further judgment based on totality of evidence. Judgment of causal association is neither simple nor straightforward, and various sets of guidelines have been proposed for the judgment. In 1965, the British statistician Sir Austin Bradford Hill proposed a list of nine guidelines to evaluate causal association, which has been widely used and is certainly the best known set of criteria for the considerations of causation, sometimes with modifications (Table 9.3).

9.4.2.1 Temporal Relationship

In causal inference, temporal relationship is essential: the cause must precede the effect; or, in other words, an exposure to cause a disease must precede the development of the disease. Of all Hill's guidelines, this is an absolute requirement. Different study designs have different strengths to provide evidence of temporal relationships. Cohort and experimental studies have obvious temporal relationships because they are performed prospectively. However, in cross-sectional studies, difficulty may arise in judging temporal relationships because the proposed cause and effect are measured at the same time point. In case-control studies, sometimes it is assumed that one event precedes another without actually establishing the order, and in other cases, it may be difficult to determine which the first is. Nevertheless, there are some strategies to find evidence supporting temporal relationships. For example, when the cause is an exposure that can be divided into different levels, it is essential that a sufficiently high level should be reached before the disease occurs. Repeated measurement of exposure at multiple time points or in different locations

Table 9.3 Hill's criteria for causation

Criteria	Comments
Temporality	The cause precedes the effect (essential criterion)
Strength	The strength of association between the cause and effect (odds ratio, relative risk)
Dose-response	Increased exposure is associated with increased effect
Consistency	Similar results are shown in other studies
Biologic Plausibility	The association is consistent with biologic mechanism
Reversibility	Removal or reduction of exposure is followed by decreased effect
Specificity	One cause leads to one effect, and vice versa
Analogy	Exposure and effect are similar to those in a well-established causal association
Experiment	Evidence from animal, intervention or mechanism studies

may also strengthen the evidence of temporal relationships. Although temporal relationships are necessary for causal inference, an existing temporal sequence alone is weak evidence for causation. Many things occur before an event: however, they have no relationship with the event. For example, someone may sneeze in Beijing city, and 30 minutes later, there's a heavy rain fall in Nanjing city. Obviously, there is no causal link between these two events.

9.4.2.2 Strength of Association

The stronger an association is, which is usually expressed by the relative risks (odds ratio, OR; relative risk, RR), the more likely a causal association is. A strong association less likely comes from either bias or confounding. Thus, the 13.70-fold higher risk of lung cancer among male smokers compared with nonsmokers is much stronger evidence than the finding that smoking is related to coronary heart disease, for which RR is only 2.00. What is a "weak' or "strong" association? There is no universal standard, but epidemiologists generally consider a relative risk (OR, RR) greater than 2.0 (or less than 0.5) to be moderately strong and a risk greater than 5.0 (or less than 0.2) to be strong. Nevertheless, a weak association does not mean that it can be overlooked for causal inference. Sometimes the strength of an association may depend on the prevalence of other possible causes. For example, the relationship between diet and coronary heart disease is a cause-and-effect association; however, the diets in populations are rather homogeneous, although greater variation may be observed among different individuals or in different stages of one person. In addition, a weak association, when combined with other guidelines, for example, consistently observed in different designs or in different settings, may also provide stronger evidence than a strong association that is only found in one or two studies.

9.4.2.3 Dose-Response Relationship

When changes in the level of possible cause are associated with corresponding changes in the incidence or prevalence of the disease, a dose-response relationship exists. Generally, the presence of a dose-response relationship in unbiased studies is considered strong evidence for causation. However, the absence of a dose-response relationship does not mean that the association is noncausal, because not all causal associations exhibit a dose-response relationship. For instance, there may be a "threshold" effect in which any exposure above a certain level will lead to disease. In addition, although a dose-response relationship is a strong evidence for causation, it cannot exclude confounding factors.

9.4.2.4 Consistency

Consistency means that, when several studies are conducted at different times in different settings or among different patient groups, the same or similar results are derived. Consistency is a kind of evidence strengthened for causal inference because the possibility that all different studies make the same "mistake" is minimized. However, a lack of consistency does not exclude a causal association. Different results may come from variation in study design or quality or different exposure levels and other conditions that may affect the impact of a causal factor on the effect.

9.4.2.5 Biologic Plausibility

A causal association generally should have biological rationality. The association between cause and effect is consistent with the current biologic knowledge and often enhances convincing causal inference. However, the lack of biological plausibility does not deny a causal association, which may simply reflect a lack of scientific knowledge or evidence. Increasing knowledge of biological mechanisms may support this association in the future. In other words, biologic plausibility, when present, enhances evidence for causation; when absent, other evidence for causation should be sought.

9.4.2.6 Reversibility

When the removal of a factor that is likely to be a cause of disease results in a decreased risk of disease, there is a greater possibility that the association is causal. An example is that people giving up smoking decreases their risk of lung cancer compared with people who continue to smoke. Reversible associations are strong but not infallible evidence for causation because they cannot exclude confounding factors, which can also conceivably account for a reversible association.

9.4.2.7 Specificity

Specificity refers to the strict corresponding relationship between a cause and a disease: that is, a certain factor can only cause one certain disease, and vice versa, the disease is just caused by a certain factor. This guideline is currently only applicable for some acute communicable or genetic diseases because for most diseases, either communicable or noncommunicable, there are many risk factors for the same effect or many diseases come from one factor. Therefore, specificity is considered the weakest evidence of all the guidelines for causation.

9.4.2.8 Analogy

Analogy is sometimes used in causal inference. Suppose there is a well-established cause-and-effect association: for instance, factor A leads to effect B. If a similar association is observed between factor C and effect D, which are also similar to factor A and effect B, respectively, we can consider that factor C is likely to be the cause of effect D. In general, analogy is weak evidence for causation.

9.4.2.9 Experimental Evidence

Experimental data, from studies in animals or other organisms, from intervention studies in humans, or from mechanistic studies, may also provide evidence supporting causal associations. In medicine, evidence from a well-conducted experimental clinical trial is always considered the strongest evidence for causation. However, in epidemiology, the results from a single or few experiments are generally not considered to be convincing or strong evidence for causation.

Among the nine guidelines described above, temporal relationships, and strength of association are necessary conditions to judge causality. That means that if there is a causal association, temporal relationships and statistically significant associations must exist; otherwise, causal associations can be denied. The other seven guidelines belong to unnecessary conditions, which are just general criteria for causal inference. A lack of any one or even all seven guidelines does not preclude causal association. Moreover, it is worth mentioning that all nine guidelines are not sufficient conditions for causation. Thus, even though a relationship between two events satisfies all nine guidelines, we cannot absolutely draw a conclusion that the relationship is causal. Causal inference is always tentative, and judgment must be made on the basis of the available evidence. Although there are no completely reliable criteria for determining whether an association is causal or not, Hill's criteria are widely accepted and have been applied in practice. Nevertheless, Hill's criteria are essentially guidelines for causal inference but not a "gold standard" to judge a causal association.

9.4.3 An Example of Causal Inference Using Hill's Criteria

Primary hepatocellular carcinoma (HCC) is one of the most common malignancies worldwide. Many researchers have investigated the causes of HCC and indicated that alcohol abuse (habitual heavy drinking) is likely to be a risk factor for HCC. The Causal inference was performed and summarized as follows.

9.4.3.1 Temporality of Association

Cohort studies have indicated that habitual heavy drinking always precedes the occurrence of HCC. Some cases of HCC had several years or even several decades of heavy drinking history.

9.4.3.2 Strength of Association

Many case-control studies have indicated that the risk (OR) of HCC among habitual heavy drinkers is 2–4-fold greater than that among nonhabitual drinkers or non-drinkers. Furthermore, a few cohort studies found that the relative risk (RR) of heavy drinking was 2–5-fold higher than that of nondrinking groups.

9.4.3.3 Consistency

The association between habitual heavy drinking and HCC has been repetitively studied in many different countries by different study designs at different times. The results from different studies consistently indicated that heavy drinking is associated with the occurrence of HCC.

9.4.3.4 Dose-Response Relationship

Previous studies have found that more alcohol consumption and a longer heavy drinking history result in a higher incidence of HCC. This is an obvious dose-response relationship.

9.4.3.5 Biologic Plausibility

The relationship between alcohol abuse and HCC is consistent with the current understanding of alcohol metabolism in the liver. Alcohol is metabolized in the liver, and its metabolism produces free radicals that cause lipid peroxidation, damage mitochondria in liver cells, and lead to alcoholic liver injury.

9.4.3.6 Experimental Evidence

A number of animal studies have shown a similar relationship between alcohol administration and hepatic injury.

Taken together, the evidence is strong for the conclusion that habitual heavy drinking is a risk factor for HCC.