

## Chapter 9

# Association and Causation

### 9.1 Introduction

The most outstanding contribution of epidemiology is the study of association and causation in health and disease. Ironically, this is also the most difficult field in epidemiology, since it is often not easy to tell whether an observed association between a condition and a risk factor represents a cause-and-effect relationship.

The reasons for interest in establishing or excluding causality are:

- to understand the determinants of disease occurrence, distribution and outcome;
- to identify the links in the chain of causality that are amenable to intervention through general or specific intervention programmes; and
- to relate the output and impact of intervention programmes to their input, i.e. a causal evaluation.

### 9.2 Defining an association

An association is said to exist between two variables when a change in one variable parallels or coincides with a change in another. This is also called 'covariation' or 'correlation'. An association or covariation may be positive or negative and may be proportionate or disproportionate. An association is said to be causal when it can be proved that a change in the independent variable (exposure) produces (induces, results in, leads to, determines or causes) a change in the

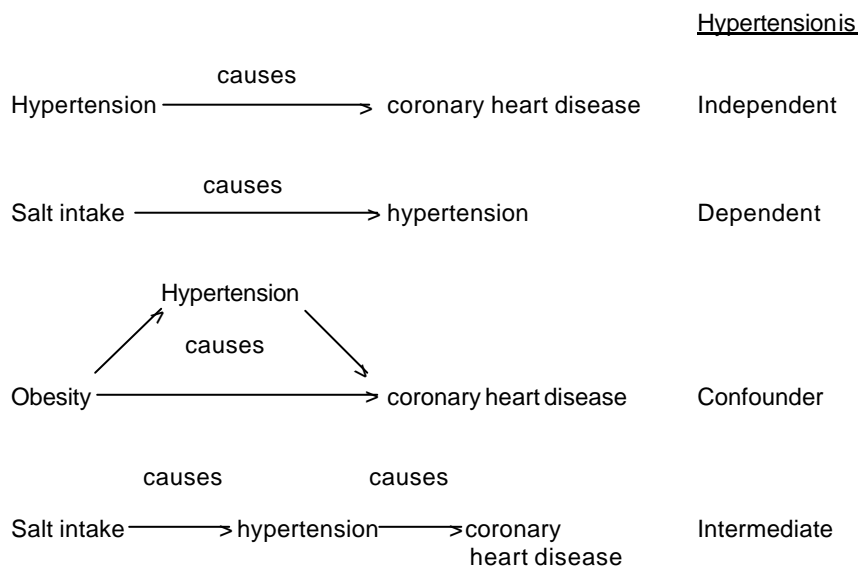
dependent variable (disease). More appropriately, a causal relationship exists when exposure enters into the causation of disease. This underlines the possibility of multiple causes.

### 9.3 Defining the variables in an association

#### 9.3.1. Independent and dependent variables

The hypothesis to be tested in a study usually defines which variable is assumed to be causal (i.e. is a risk factor) and which variable is considered to be the effect. The definition of a variable therefore depends on the study hypothesis: a variable may be independent in one hypothesis, a confounder in another, and dependent in a third. Take for instance, 'hypertension' in the simplified models shown in Figure 9.1.

FIGURE 9.1 VARIABLES INVOLVED IN HYPERTENSION



### **9.3.2. Confounding variables**

A confounding variable is an independent variable (other than the hypothesized causal variable) that has or can have an effect on the dependent variable, but the distribution of which is systematically correlated with that of the hypothesized causal variable.

### **9.3.3. Control variables**

Control variables are independent variables (other than the causal variable) which are potential confounders, and hence should be controlled or neutralized in the design or analysis. Note that these are only the 'known' or controllable variables. In most studies, it is impossible to control for all variables other than the suspected causal variables. By the process of randomization, it is hoped that many of these 'uncontrolled' variables will be equally distributed between the exposure and control groups.

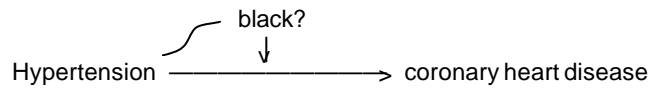
### **9.3.4. Intermediate or intervening variables**

When the effect of a causal variable on the dependent variable or study condition is mediated through a third set of variables, the latter are called intermediate variables. They are in fact dependent variables in relation to the causal variable, and independent in relation to the subsequent condition. Recall the example in Figure 9.1, in which the effect of salt on coronary heart disease was mediated through hypertension. The role of intermediate variables must be given more emphasis in epidemiology, especially when the hypothesized causal variables are global, such as, 'social condition' or 'development', when their association with, e.g. infant mortality is considered. In answering the question, 'How, in fact, does an increase in income or education bring about a reduction in infant mortality?' we should, for example, consider increased access to prenatal care, better maternal and infantile nutrition, access to vaccination, and better housing and personal hygiene. These are intermediate variables, some of which should be specified in the study design, and about which data are collected.

### **9.3.5. Effect modification**

Some independent variables may modify (positively or negatively) the effect of the hypothesized causal variables. For example, hypertension is more frequent among black than among white Americans, while coronary heart disease is more frequent in whites

than in blacks. It is possible, therefore, that something related to the constitution or way of life of blacks modifies the effect of hypertension on coronary heart disease among them. Thus,



Some confounding variables are also effect modifiers.

#### 9.4 Measuring an association

When the incidence (or prevalence) of a condition (e.g. lung cancer) in a group with certain characteristic (e.g. smoking) differs from the incidence (or prevalence) in a group without the characteristic (e.g. non-smokers), an association is inferred that may or may not be causal. The strength of the association is commonly measured by the relative risk or odds ratio (OR), in addition to attributable risk and population attributable risk per cent.

Another measure of association is the correlation between two variables. This can be expressed graphically in a correlation or scatter diagram (Figure 9.2), when the dependent variable (e.g. lung cancer incidence or mortality) is plotted on the vertical or Y-axis and the independent variable or characteristic (e.g. number of cigarettes smoked) on the horizontal or X-axis. If an association exists, changes in Y will coincide with changes in X. The relationship can also be expressed in terms of a correlation coefficient, or  $r$ , which is a measure of the degree to which a dependent variable varies with an independent variable. The correlation coefficient varies between +1 and -1. Table 9.1 gives approximate degrees of association corresponding to levels of  $r$ , subject, of course, to statistical tests of significance.

**TABLE 9.1 APPROXIMATE DEGREE OF ASSOCIATION  
CORRESPONDING TO LEVEL OF  $r$**

$r$	Degree of association
$\pm 1.0$	Perfect
$\pm 0.7$ to $\pm 1.0$	Strong
$\pm 0.4$ to $\pm 0.7$	Moderate
$\pm 0.2$ to $\pm 0.4$	Weak
$\pm 0.01$ to $\pm 0.2$	Negligible
0.0	No association

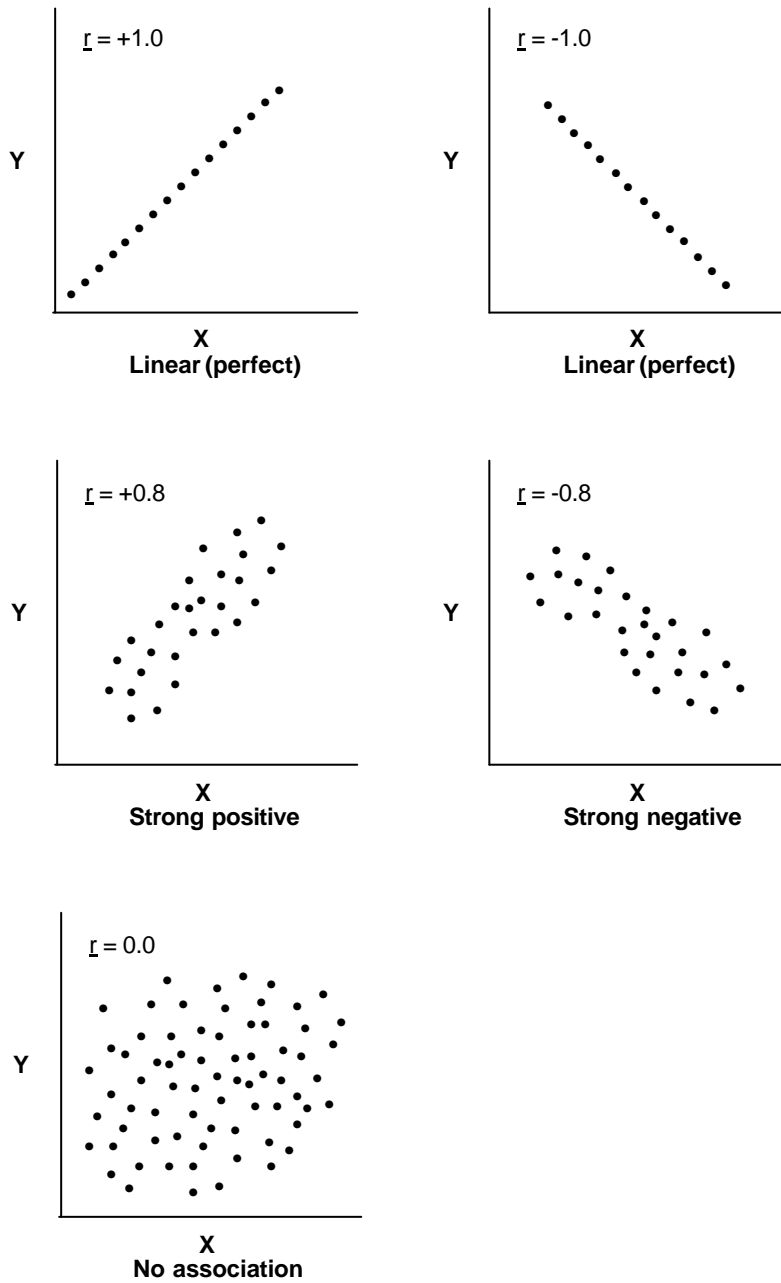
The common correlation coefficients in use include:

- the Pearson product-moment correlation coefficient,
- the Spearman rank-order correlation coefficient, and
- the Kendall tau correlation coefficient.

Regression coefficients can also be used in measuring association. They are a measure of the mean changes to be expected in the dependent variable for a unit change in the value of the independent variable.

When more than one independent variable is associated with the dependent variable, multiple regression analysis will indicate how much of the variation observed in the dependent variable can be accounted for by one, or a combination of independent variables.

FIGURE 9.2 SCATTER DIAGRAMS AND CORRELATION COEFFICIENTS ( $r$ )



## **9.5 Problems in establishing causality**

1. The existence of a correlation or association does not necessarily imply causation.
2. The concept of a single cause (the agent), once held in relation to communicable disease, has been replaced by the concept of multiple causation in diseases such as cancer and heart disease. Even in communicable diseases, factors in the agent, the host and the environment cooperate to cause the disease. For example, the tubercle bacillus is a necessary, but not a sufficient factor in the causation of tuberculosis.
3. The criteria used in establishing causality in infectious disease, namely, Koch's postulates, are not applicable to non-infectious diseases. Koch's postulates are:
  - a. The organism is always found with the disease in accord with the lesions and clinical stage.
  - b. The organism is not found in any other disease.
  - c. The organism is isolated from one who has the disease, and cultured through several generations.
  - d. The organism from culture is capable of producing disease in susceptible animals.

Even in some infectious diseases, these postulates are not totally applicable.

4. The period between exposure to a factor or cause, and the appearance of clinical disease, is relatively long in non-infectious diseases. During this latent period, exposure to other factors complicates the research.
5. Specificity, easily established in infectious disease, does not apply to most other diseases. Lung cancer, for example, can result from smoking or exposure to radiation, asbestos or nickel dust. At the same time, each of these risk factors can cause diseases other than lung cancer. Smoking, for example, is involved in the causation of heart disease and emphysema.
6. Certain 'noise' factors, or confounders that are associated with the cause of a disease tend to distort or confound the relationship with the suspected factors. These require special handling during design or analysis to control or neutralize their effect.

7. Several systematic errors or bias in research design or data collection can produce false or spurious associations.
8. No statistical method can differentiate between causal and non-causal associations.

Because of these many uncertainties, the terms 'causal inference', 'causal possibility' or 'likelihood' are preferred to 'causal conclusion'. Such inferences would be enough in many situations to formulate policy rather than waiting for the unequivocal proof, which may be unattainable in several disease conditions.

### 9.6 Steps in establishing causality

Epidemiological strategies are usually assessed according to their power to provide a basis for causal inferences. It should be emphasized, however, that causal inference should not be made until certain requirements have been satisfied, which relate to two major questions:

- Is there actually an association?
- If there is an association, is it likely to be causal?

The requirements for making a causal inference aim (i) to exclude a non-causal association, and (ii) to ascertain the likelihood of a causal association. The requirements are given below:

1. The association actually exists and is statistically meaningful.

This requires that:

- a. The association is not due to chance, as asserted by statistical tests of significance that can be applied to the difference between the frequency of the disease (the dependent variable) among those with and those without exposure to the risk factor (the independent variable). Tests can also be applied to the relative risk of disease in the two groups or to the correlation coefficient. Such tests would determine how frequently an association of the observed magnitude would occur solely on the basis of random variation or chance.

- b. The association exists at the individual level and is not based only on the association measured on an ecological level, i.e. when the aggregate or geographical unit was used as the unit of observation. The possibility of ecological fallacy precludes inferring causality on an individual level.
  - c. The association is not based on numerator analysis, i.e. per cent distribution of 'cases' (the dependent variable), but on the appropriate population-based rates, calculating the relative risk or odds ratio.
2. The association is not spurious (i.e. not due to bias).
- Spurious association can be of three types:
- due to selection bias,
  - due to information or measurement bias, and
  - due to confounding bias.
- This will be discussed in detail below.
3. The confirmatory criteria for causality are satisfied.

Even if a statistical association does not exist and is not due to bias, a causal inference cannot be made confidently without satisfying the confirmatory criteria of causality. These relate to specific qualities of the association between the risk factor and the disease, namely, its strength, biological gradient, temporality, coherence, biological plausibility, specificity, consistency and experimental proof. These criteria are elaborated upon below.

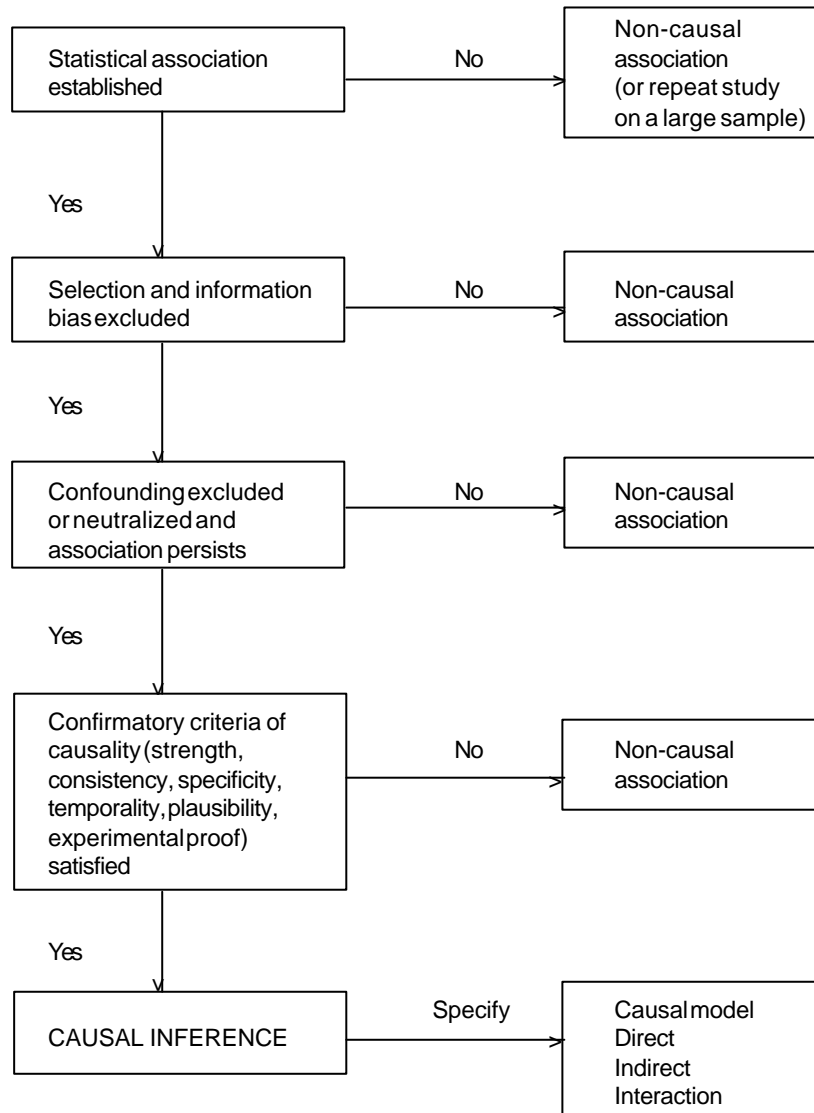
The steps for establishing causality are represented diagrammatically in Figure 9.3.

Note: When intervening variables or mechanisms are involved, information on these variables should be collected as well.

## 9.7 Confirmatory criteria for a causal inference

Having established a statistical association and having ruled out sources of bias (i.e. having established that the association is not spurious), other specific criteria should be satisfied to support the causal inference.

FIGURE 9.3 - ESTABLISHING A CAUSAL INFERENCE



***The association is strong (strength).***

The strength of the association is measured by the relative risk (and attributable risk) and OR (in case-control studies). Correlation and regression coefficients can endorse these measures of effect. The stronger the association, the higher the likelihood of a causal relationship.

***There is biological gradient.***

A dose-response relationship (if present) can increase the likelihood of a causal association. This is not, however, possible in all studies.

***The association follows a time sequence (temporality).***

It goes without saying that the risk factor or cause must precede the condition or effect. This antecedent-consequence requirement is often overlooked. It is easier to establish temporality in experimental and cohort studies than in case-control and cross-sectional studies.

***The association is plausible (coherence or plausibility).***

The association should make common biological or sociological sense and should not conflict with existing theories or knowledge unless it is actually a challenge to those theories. In either case, there should be some theoretical basis explaining the association.

***The association is consistent (consistency).***

Causality is more likely when the association is supported by other investigations conducted by different persons in different places, circumstances and time-frames, and using different research designs.

***The association is specific (specificity).***

The disease outcome should be specific to, or characteristic of, exposure to a particular risk factor. This is more feasible in infectious diseases than in non-infectious diseases, which can result from different risk agents. Hence, this criterion is not generalized.

***There is experimental proof for causality.***

Two types of experimental proof can be established: (i) experiments in humans using the risk factor, which are difficult to provide, and (ii) cessation experiments, whereby removal of the putative cause results in a significant reduction in disease incidence.

## 9.8 Types of association

The association between two variables may be causal or non-causal.

### 9.8.1. Causal association

As already stated, a causal association exists when the independent variable (risk factor) causes changes in the dependent variable. Causal associations are of three types (see Figure 9.4).

a. Direct causal association

A direct causal association is inferred when the risk factor or independent variable changes the dependent variable or condition directly, without intervening variables, e.g. exposure to the tubercle bacillus causes tuberculosis, exposure to lead causes lead poisoning, and iodine deficiency causes goitre.

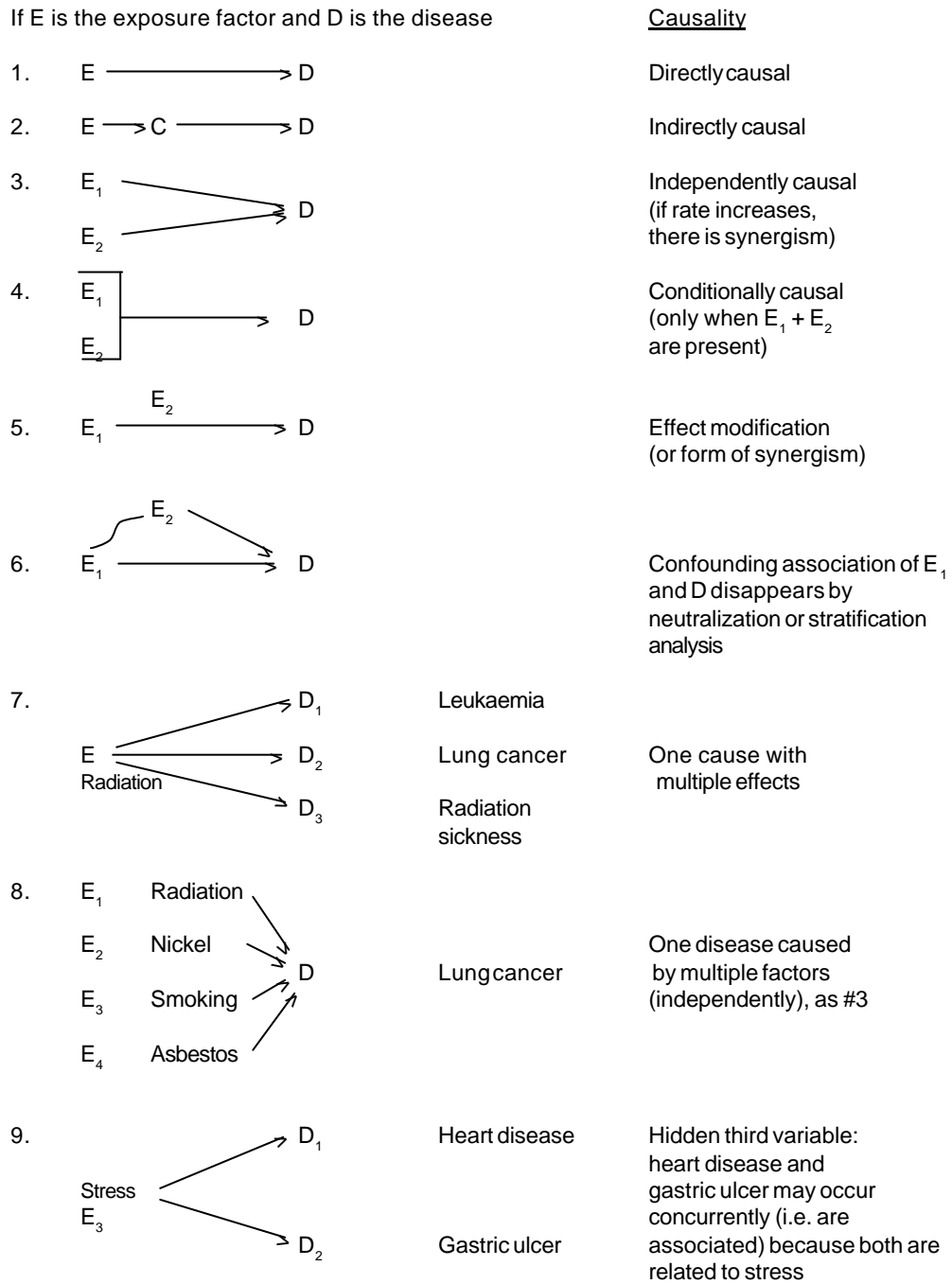
b. Indirect causal association

The association is inferred when the risk factor or independent variable causes changes in the dependent variable or condition through the mediation of other intermediate variables or conditions:

iodine deficiency —————> goitre —————> thyroid adenoma

Thus, thyroid adenoma is caused indirectly by iodine deficiency. Note that the term 'indirect association' may be used in a broader sense. For example, endemic goitre is associated with high altitude simply because water supplies are likely to contain less iodine at high rather than low altitudes. Such usage, however, should be restricted and carefully evaluated. The main issue here is whether the association is causal or non-causal. The criteria for causality should apply equally to direct and indirect causal associations.

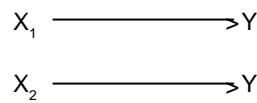
FIGURE 9.4 CAUSAL MODELS



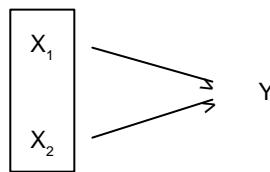
c. Interaction (including conditional) causal association

There may be interactions (positive or negative) between categories of independent variables that produce changes in the dependent variables.

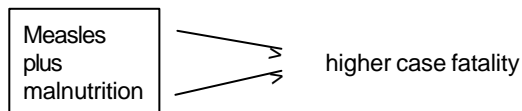
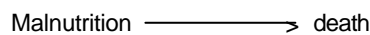
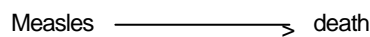
One form is synergism (or antagonism) between two variables, whereby each factor has an independent effect on the condition, while the joint effect is greater (or smaller) than each alone. In one-way analysis, each factor has an effect on the condition:



In stratification analyses (e.g. control table analysis), neither effect disappears, but the joint effect may be greater (or smaller):



For example, measles can result in death, but the probability is greater in malnourished children:

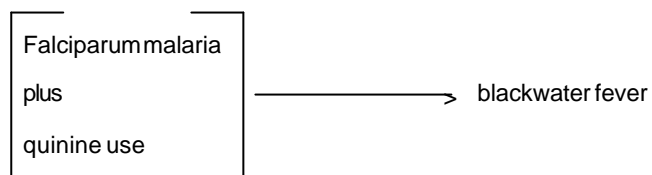


In a conditional causal association, two risk factors are incapable of producing a condition unless they exist in the presence of each other. For example, blackwater fever (a febrile condition characterized

by dark urine due to haemolysis) follows malaria as a complication only if the malaria was due to *Plasmodium falciparum* and the cases were treated with quinine.

Falciparum malaria alone —————> no blackwater fever

Quinine alone —————> no blackwater fever



### 9.8.2. *Non-causal, spurious association*

In some situations, an association does exist, but, despite its significance and strength, it may be spurious or non-causal as far as the special characteristics under study are concerned. A non-causal association is inferred when the association is:

- due to chance,
- based on numerator analysis or ecological correlation, or
- due to bias.

## 9.9 References and further reading

Hill A.B. Principles of medical statistics. New York, Oxford University Press, 1966.

Kelsey L.T., Thompson W.D., Evans S.A. Methods in observational epidemiology. New York, Oxford University Press, 1986.

Kleinbaum D.G., Kupper L.L., Morganstern H. Epidemiologic research: principles and quantitative methods. New York, Van Nostrand Reinhold, 1982.

Lilienfeld A.M., Lilienfeld D.E. Foundations of epidemiology, 2 ed. New York, Oxford University Press, 1980.

[<< Back to Table of Contents](#)