
Chapter 5

Overview of study designs

In epidemiology, measuring the occurrence of disease or other health-related events in a population is only a beginning. Epidemiologists are also interested in assessing whether an exposure is associated with a particular disease (or other outcome of interest). For instance, researchers may be interested in obtaining answers to the following questions:

- * *Does a high-fat diet increase the risk of breast cancer?*



- * *Does hepatitis B virus infection increase the risk of liver cancer?*



The first step in an epidemiological study is to define the hypothesis to be tested. This should include a precise definition of the exposure(s) and outcome(s) under study. The next step is to decide which study design will be the most appropriate to test that specific study hypothesis.

5.1 Types of study design

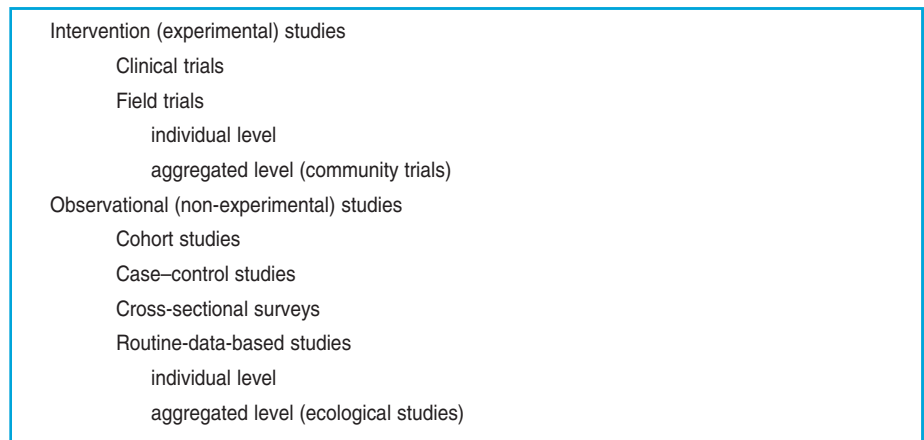
There are two basic approaches to assessing whether an exposure is associated with a particular outcome: experimental and observational. The *experimental* approach is perhaps more familiar to clinicians, since it corresponds to the approach used for investigations in laboratory-based research. In an experiment, investigators study the impact of varying some factor which they can control. For example, the investigators may take a litter of rats, and randomly select half of them to be exposed to a supposedly carcinogenic agent, then record the frequency with which cancer develops in each group. The equivalent of this animal experiment in human beings would entail selecting a group of individuals, randomly allocating half of them to exposure to a hypothesized disease-producing factor, and then comparing the occurrence of disease in subjects who were exposed with that in subjects who were not. For ethical reasons, it would be impossible to conduct such a study in human subjects. It is, however, possible to conduct a trial to test whether removal of such an exposure will decrease subsequent incidence and mortality. Thus, experimental studies in epidemiology are limited to interventions that are believed to be of potential benefit.

Since epidemiologists can rarely conduct experiments, their role is usually limited to observing the occurrence of disease in people who are already segregated into groups. For instance, we can follow up people who happen to be or not be infected with the hepatitis B virus, to see whether their risks of liver cancer are different. These studies are *observational*, because the role of the investigator is merely to observe what happens, noting who is exposed or unexposed and who has or has not developed the outcome of interest.

A major problem with observational studies is that the observed groups may differ in many other characteristics, in addition to the one under study. Thus, people in various occupations may differ not only in exposure to occupational hazards but also in other lifestyle characteristics such as socioeconomic background, health status, fitness for the job, smoking and alcohol habits and many other factors. Because of these *confounding* and often unmeasurable factors, the role of a specific exposure under investigation is more difficult to establish than in experimental studies.

Epidemiological studies can thus generally be classified as *intervention* or *observational* studies. Within each of these two broad categories, studies can be further organized as in [Figure 5.1](#).

Figure 5.1.
Main types of epidemiological study design.



The classification scheme shown in [Figure 5.1](#) should be taken as just a simple way of presenting and discussing the different study designs. There are, of course, many other ways in which epidemiological study designs may be classified. Also, in reality, many studies have mixed features.

5.1.1 Intervention studies

Intervention studies are characterized by the fact that the study subjects are allocated by the investigator to the different study groups through the use of *randomization* ([Figure 5.2](#)). This ensures that the assignment of subjects to the various groups is determined by chance alone and is not subjectively influenced by the investigators or the participants (see [Section 7.9.1](#)). There are two main types of intervention study: clinical trials and field trials.

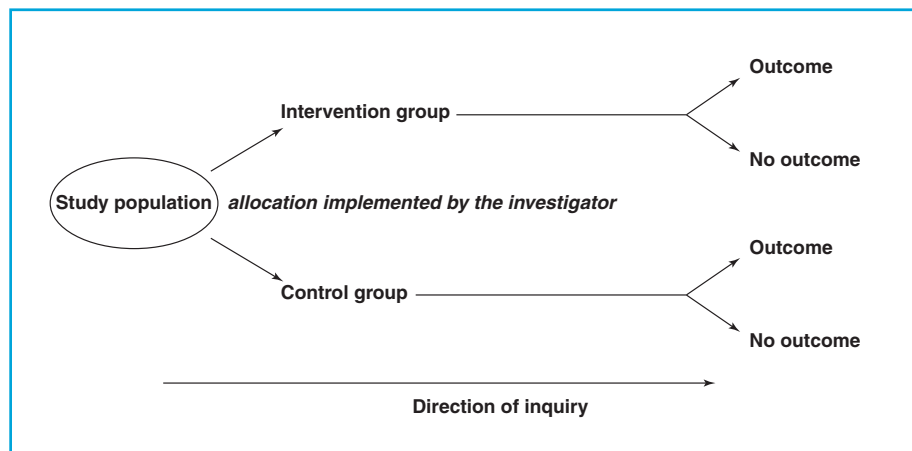


Figure 5.2.
Outline of an intervention trial.

The objective of a *clinical trial* is to evaluate new forms of treatment of a disease or condition. Thus clinical trials are usually carried out in hospitals or clinics among people who have already developed the disease.

Example 5.1. *The Women's Intervention Nutrition Study was set up to assess whether a low-fat diet will reduce cancer recurrence and improve survival of women who have surgery for early and moderate stage breast cancer. A total of 2000 postmenopausal women with resected breast cancer are being recruited and randomized to receive standard treatment plus a low-fat dietary intervention or standard treatment alone. The standard treatment consists of tamoxifen, and radiotherapy and/or chemotherapy as appropriate. The study will last for five years (Chlebowski & Grosvenor, 1994; Henderson, 1995).*

In [Example 5.1](#), women who had already developed breast cancer (and had surgery) were randomized to standard treatment accompanied by a low-fat dietary regimen ('intervention group') or standard treatment alone ('control group') to assess whether a low-fat diet reduces recurrence of the tumour and increases survival.

Example 5.2. *The Women's Health Initiative is a randomized trial taking place in the USA to determine whether a sustained low-fat diet will reduce the incidence of breast cancer. A total of 48 000 postmenopausal women with no prior history of breast or colon cancer are being randomized to the intervention or control group. The dietary intervention is designed to reduce fat intake to 20% of total kilocalories and to increase intake of fruits and vegetables. The trial will last 11 years (Chlebowski & Grosvenor, 1994; Henderson, 1995).*

In contrast to clinical trials, *field trials* deal with subjects who are disease-free and, therefore they generally have to be conducted in the 'field' rather than in hospitals or clinics. Their main objective is to evaluate whether an agent or procedure reduces the risk of developing disease among those free from the condition at enrolment.

In [Example 5.2](#), women are being randomized to receive ('intervention group') or not receive ('control group') the dietary intervention programme.

Whereas complications from breast cancer are quite common among women who already suffer from the disease, the risk of healthy women developing breast cancer is very small, even in high-risk populations. As a result, the field trial in [Example 5.2](#) involves a much larger number of subjects followed for a much longer period than the clinical trial in [Example 5.1](#).

Example 5.3. *The Gambia Hepatitis Intervention Study is a large-scale hepatitis B vaccination project initiated in The Gambia in July 1986. In this trial, 60 000 infants received a course of hepatitis B vaccine and a similar number did not. A national surveillance system will detect all new cases of hepatocellular carcinoma (and other chronic liver diseases) over a period of 30 to 40 years (Gambia Hepatitis Study Group, 1987).*

In [Example 5.3](#), half the infants participating in the trial received a course of hepatitis B vaccine ('intervention group') and the other half did not ('control group'). This study has a remarkably long follow-up period (it will not be able to answer the main research question before the year 2020!). This is because the intervention had to be given to infants before they were infected with hepatitis B, whereas most cases of hepatocellular carcinoma are expected in adulthood.

These last two examples clearly illustrate the complexities involved in the design and implementation of field trials when cancer is the outcome of interest. Because cancer is a relatively rare condition, they require the enrolment of large numbers of subjects who have to be followed up for relatively long periods of time. One advantage of such trials is that it is possible to assess the impact of the intervention on several outcomes. For instance, the effect of a low-fat diet on the incidence of colon cancer ([Example 5.2](#)) and of hepatitis B vaccination on the incidence of other chronic liver diseases ([Example 5.3](#)) are also being assessed in the trials described.

In field trials, the unit of allocation to the intervention may be the *individual* (as in [Example 5.2](#)) or, alternatively, a *group* of people. The group may be a household, a school or a whole community. If the unit of allocation is a community (as in [Example 5.4](#)) the study is called a community trial.

Intervention trials are very powerful for testing hypotheses. Despite this, they are not the most commonly used study design in epidemiology, mainly because of ethical constraints.

Example 5.4. *The Community Intervention Trial for Smoking Cessation is an on-going trial designed to evaluate a community-wide smoking cessation programme in the USA. Eleven pairs of communities were selected and one member of each pair was randomly assigned to receive a smoking cessation programme, while the other acted as a control. The intervention was designed to promote smoking cessation by using a wide range of community resources to affect community attitudes and policies towards smoking (COMMIT Research Group, 1991).*

5.1.2 Cohort studies

Cohort studies are observational studies in which the starting point is the selection of a study population, or cohort. Information is obtained to determine which members of this cohort are exposed to the factor of interest. The entire population is then followed up over time and the incidence of the disease in the exposed individuals is compared with the incidence in those not exposed (Figure 5.3). This type of observational study is the one that most closely resembles intervention studies, except that allocation of subjects to the exposure is *not* controlled by the investigator.

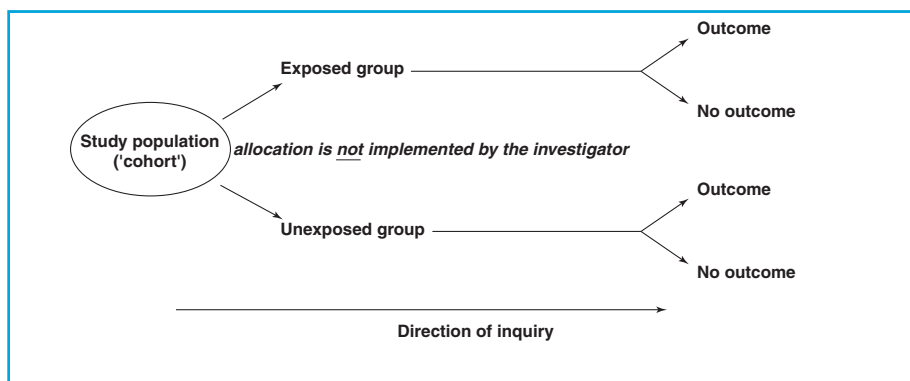


Figure 5.3. Outline of a cohort study.

Example 5.5. *In 1980, 89 538 registered nurses in the USA, aged 34 to 59 years and with no past history of cancer, completed a previously validated dietary questionnaire designed to measure individual consumption of total fat, saturated fat, linoleic acid and cholesterol, as well as other nutrients. The nurses were then classified in five groups of similar size according to their levels of fat intake, followed up in time and the incidence of breast cancer in each of these groups measured and compared (Willett et al., 1987).*

Cohort studies take individuals and classify them according to their exposure status. Sometimes they can be simply classified as exposed/unexposed. More usually, various degrees of exposure can be identified. In Example 5.5, the US nurses (the 'cohort') were classified into five groups (quintiles) according to

their levels of fat intake. The incidence of breast cancer ('outcome') was then measured and compared across these quintiles.

Example 5.6. A cohort study of Chinese male government employees in Taiwan was set up to investigate the association between hepatitis B virus infection and the development of primary hepatocellular carcinoma. All participants completed a health questionnaire and provided a blood sample at their entry into the study. A total of 22 707 men were enrolled into the study: 3454 were positive for hepatitis B surface antigen (HBsAg) and 19 253 were negative. These men were then followed up and the incidence of hepatocellular carcinoma among HBsAg carriers was compared with the incidence among non-carriers (Beasley et al., 1981).

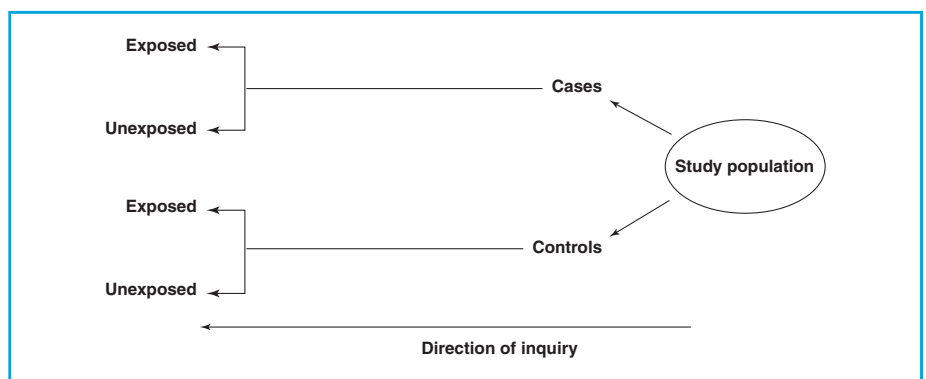
In [Example 5.6](#), the cohort consisted of Chinese male government employees in Taiwan. Subjects were classified as HBsAg carriers ('exposed') or non-carriers ('unexposed') at the time of their enrolment into the study and were followed up to assess whether the risk of hepatocellular carcinoma ('outcome') was higher in those exposed than in those unexposed.

In cohort studies, it is important that the groups being compared are as similar as possible with respect to all other factors that may be related to the disease. Since the investigator has no control over who is or is not exposed, it is likely that the exposure groups will differ in relation to factors other than the one being investigated, so that special techniques have to be used in the analysis to ensure that these uneven distributions are taken into account (see [Chapters 8, 13 and 14](#)).

5.1.3. Case-control studies

Case-control studies are observational studies in which the starting point is the identification of 'cases' of the disease (or condition) of interest, and of suitable 'controls' without that disease (or condition). Cases and controls are then compared to assess whether there were any differences in their past exposure to putative risk factors ([Figure 5.4](#)).

Figure 5.4.
Outline of a case-control study.



Example 5.7. *A case-control study was conducted in Singapore to investigate the role of diet in breast cancer. Two hundred Chinese women with histologically confirmed breast cancer and 420 controls without this disease participated in the study. A dietary questionnaire was used to measure past dietary intake. Cases and controls were then compared to assess whether there were any differences in their past intake of selected foods and nutrients (Lee et al., 1991).*

In [Example 5.7](#), women with ('cases') and without ('controls') breast cancer were identified and their past diet ('exposure') compared.

Example 5.8. *A case-control study was carried out in Taiwan to assess whether hepatitis B infection played a role in the etiology of primary hepatocellular carcinoma. A total of 128 histologically or cytologically confirmed cases of hepatocellular carcinoma and 384 controls without the disease were included in the study. Of the cases, 77% were carriers of the hepatitis B surface antigen (HbsAg) compared with only 28% of the controls (Chuang et al., 1992).*

In [Example 5.8](#), subjects with ('cases') and without ('controls') hepatocellular carcinoma were identified and the frequencies of the HBsAg carrier status ('exposure') in the two groups were compared.

The major difference between cohort and case-control methods is in the selection of study subjects. In a cohort study, subjects are selected who are initially free of disease and are then followed over time. In contrast, in the case-control approach, subjects are selected on the basis of the presence or absence of the disease (or any other outcome) under study.

Case-control studies are particularly suitable for investigating rare diseases such as cancer. A cohort study would require the follow-up of a large number of individuals for a long period of time in order to accrue enough cases of a rare disease. Case-control methods are also more appropriate for studying diseases with a long induction period. This is because the case-control study starts with subjects who have already developed the condition of interest, so there is no need to wait for time to elapse between an exposure and the manifestation of disease as in cohort studies.

The number of subjects necessary for a case-control study is much smaller than the number required for cohort studies. [Examples 5.5](#) and [5.6](#) involved large numbers of subjects followed up for relatively long periods of time. The same questions were addressed in [Examples 5.7](#) and [5.8](#) using much smaller numbers of subjects. Thus, case-control studies are relatively inexpensive to carry out, at least compared with cohort studies.

Results from case-control studies are, however, more difficult to interpret. First, controls should represent the same study population from

which the cases were drawn. If not, the results will be distorted by *selection bias* (see Chapter 9). Secondly, one cannot be sure that the exposure did precede the disease (except when the exposure is a fixed attribute such as blood type that does not change over time). In [Example 5.7](#), it is conceivable that the reported diet may be a *consequence* rather than a *cause* of breast cancer.

5.1.4 Cross-sectional surveys

In a cross-sectional survey, a sample of individuals is selected from a previously defined population and contacted at a particular point in time to obtain simultaneously information on both the exposure(s) and outcome(s) of interest.

Example 5.9. *To assess whether there is an association between Helicobacter pylori infection and chronic atrophic gastritis, a relatively common condition and an established precursor of gastric cancer, a cross-sectional survey was performed among 1815 randomly selected healthy blood donors in four prefectures of Japan. Blood samples were taken from all study subjects and measurements made of serum H. pylori IgG antibodies and serum pepsinogen I and II (markers of chronic atrophic gastritis). The prevalence of antibodies against the bacterium among subjects with chronic atrophic gastritis was then compared with the prevalence among individuals without chronic atrophic gastritis (Fukao et al., 1993).*

In [Example 5.9](#), a sample of 1815 study subjects was selected from a well defined population (healthy blood donors in certain prefectures of Japan) and a blood sample taken from each of them at a particular point in time to measure both the levels of *H. pylori* antibodies ('exposure') and the levels of pepsinogen I and II ('outcome').

In this type of study, it is crucial to ensure that the sample of subjects who participate in the study is representative of the whole population to whom the results will be extrapolated. Otherwise, the results will be distorted by *selection bias*. The best way to safeguard against selection bias is to use *random sampling* methods. These methods ensure that chance alone determines who will and who will not be included in the sample (see Chapter 10).

Cross-sectional surveys are generally used to estimate the prevalence of common conditions of reasonably long duration. Thus this design is not appropriate for studying diseases such as cancer, since it would be necessary to survey a very large population to identify enough cases to be able to draw any conclusions. Moreover, prevalent cancer cases are a biased sample of all cases, in which those with long survival tend to be over-represented.

The main use of cross-sectional surveys in cancer epidemiology has been to examine the distribution and determinants of common conditions, such as human papillomavirus infection or skin naevi, which are known (or suspected) to be associated with cancer. This type of study has also been used

to investigate the distribution and determinants of known (or potential) high-risk behaviours, such as being a smoker or being a regular sunbed user.

These studies are relatively simple to conduct and take only a short time because they do not require follow-up of the study subjects. Their main disadvantage is that, as with case-control studies, it is not possible to know whether the outcome followed the exposure in time or the exposure resulted from the outcome, since information on both exposure and outcome is collected at the same single point in time. In [Example 5.9](#), it is not possible to establish whether *H. pylori* infection preceded or followed chronic atrophic gastritis. This is obviously not a problem for exposures that do not change over time such as gender, ethnicity or genetically determined traits. For exposures that are likely to change over time, cross-sectional surveys may include questions about past as well as current exposures. For example, in a health survey of workers in a particular industry, workers may be asked details about their current job and any other jobs they have had in the past.

5.1.5. Routine-data-based studies

The distinguishing feature of this type of study is that the data on the exposure(s) and outcome(s) of interest are obtained from routine data-collection systems, without the researcher contacting any of the study subjects. Routine-data-based studies can be carried out at an *individual* or at an *aggregated* level.

Individual level

Many routine data-collection systems collect data on personal attributes such as age, sex, place of birth, place of residence, occupation, etc. Cancer occurrence can then be examined in relation to these variables. The objective is to search for patterns that might suggest or confirm specific etiological hypotheses.

In [Example 5.10](#), data on place of birth, place of residence and age at migration ('exposures') and on breast cancer ('outcome') were available from population censuses, social security records and cancer registries for each of the study individuals.

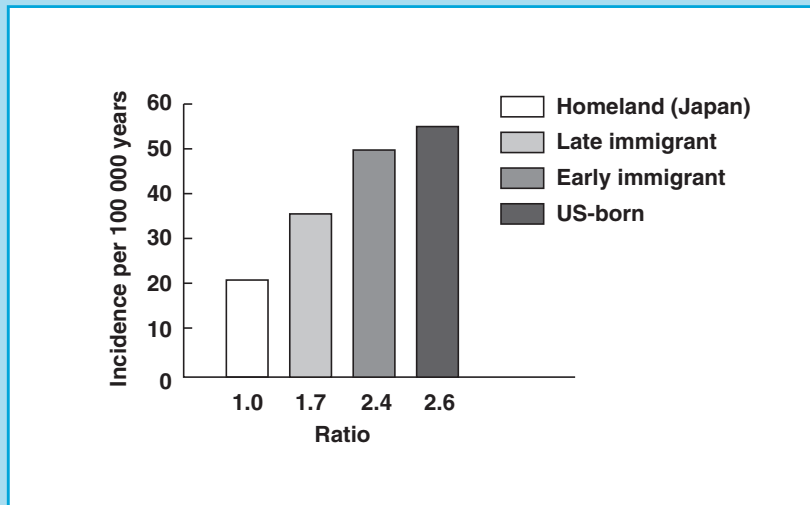
The key feature of this type of study is that data on both the exposure and the outcome(s) of interest are obtained for each of the study subjects from routine data-collection systems. In terms of their analysis and interpretation, they can be regarded as being similar to cohort or case-control studies. For instance, [Example 5.10](#) can be regarded as a cohort of women who were classified into four different exposure categories according to their place of birth, residence and age at migration. These women were then followed up in time and the occurrence of breast cancer in each group was measured and compared.

The main advantage of routine-data-based studies in relation to other observational studies is that they can be carried out relatively quickly and cheaply because the data have already been collected and there is no need to contact the individuals. A major limitation, however, is that few vari-

Example 5.10. The risk of female breast cancer was studied in four population groups: (1) Japanese women born and resident in Japan (homeland); (2) Japanese women who migrated later in life to the USA; (3) Japanese women who migrated early in life to the USA; and (4) women born in the USA. The main results are shown in Figure 5.5. They suggest that environmental factors (e.g. diet) in early life may be important in the development of breast cancer (Shimizu et al., 1991).

Figure 5.5.

Age-adjusted incidence rates of female breast cancer for USA residents (1972–85) by birthplace and age at migration, and for Japan (1973–81) (reproduced, by permission of Churchill Livingstone, from Shimizu et al., 1991).



ables are usually available and they have not been collected with the specific needs of the study in mind. In [Example 5.10](#), the variables ‘country of birth’ and ‘country of residence’ are only proxy measures for more biologically relevant exposures, such as reproductive factors, diet and other lifestyle characteristics, for which no data were available in the routine data-sets.

Aggregated level (ecological studies)

Studies which involve investigating the frequency of disease (or any other outcome of interest) in relation to the level of exposure in several groups of people (or in the same group over different periods of time) are called ecological studies. In this type of study, it is not possible to link the exposure of a particular individual to his or her outcome. Thus, the *group* rather than the individual is the unit of observation and analysis. The groups may be defined in a large number of ways, such as according to place of residence, place of birth, socioeconomic status, occupation, etc.

Most ecological studies in cancer epidemiology make use of routinely collected data. Data on the average (or frequency of) exposure in different population groups may be available from government or private sources that routinely collect data on demographic, environmental and lifestyle

variables. Disease rates may be available from surveillance programmes, cancer registries or death certification systems.

In [Example 5.11](#), it is not possible to link the breast cancer experience of any individual woman with her diet because the only two pieces of information available for each of the countries were their breast cancer incidence rate and an estimate of their average *per caput* consumption of fat.

Example 5.11. Breast cancer incidence data and average per caput daily consumption of a wide range of foods for 24 countries were extracted from routinely collected data sources (Armstrong & Mann, 1985). The relationship between fat consumption and breast cancer is illustrated in Figure 5.6.

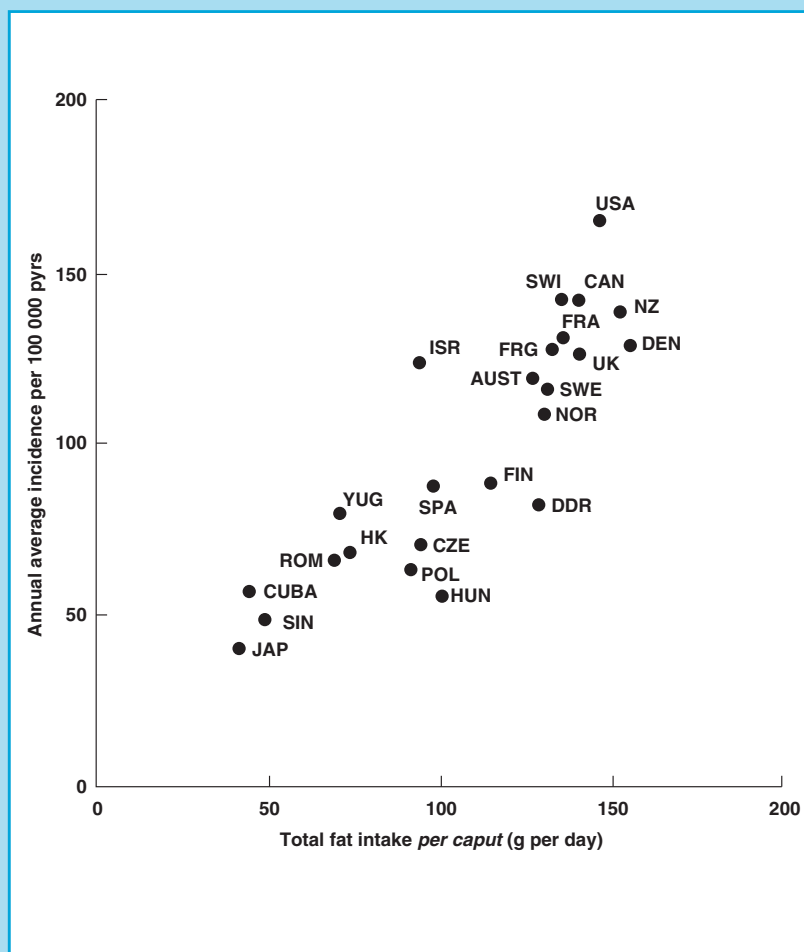


Figure 5.6.

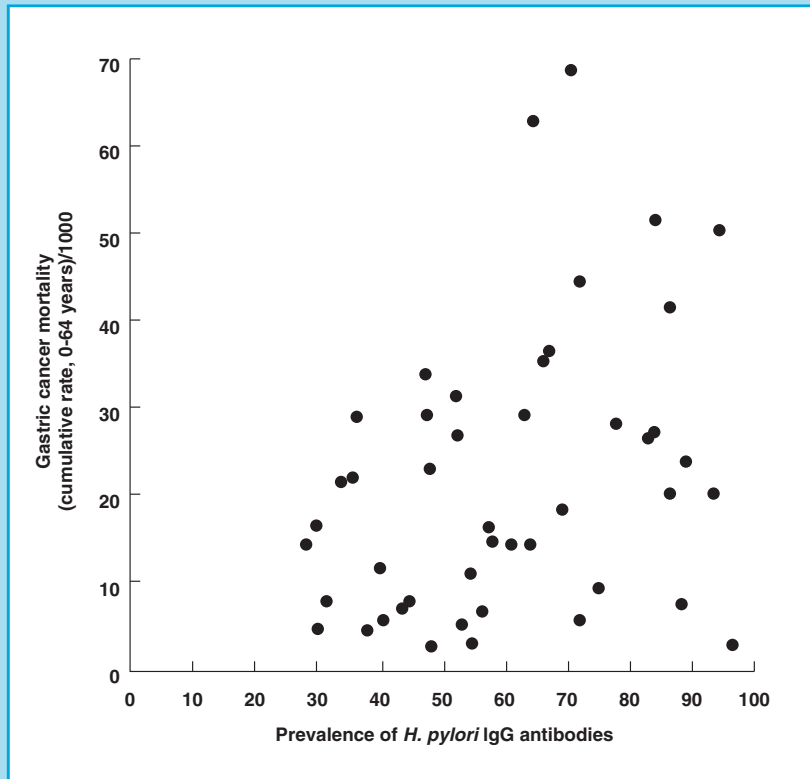
Estimated daily consumption of fat *per caput* in 1964-66 and age-adjusted breast cancer incidence rates in women aged 35-64 years in 1972-77 in 24 countries (reproduced, by permission of Oxford University Press, from Armstrong & Mann, 1985).

Sometimes data on the exposure(s) or outcome(s) of interest may not be available from routine data-collection systems but may be obtained from previously conducted surveys (as in [Example 5.12](#)).

Example 5.12. A study was carried out to examine the association between prevalence of *Helicobacter pylori* infection and mortality from gastric cancer in 46 rural counties of the People's Republic of China. The results are plotted in Figure 5.7 (Forman *et al.*, 1990).

Figure 5.7.

Prevalence of *H. pylori* IgG antibodies and gastric cancer mortality in 46 rural Chinese counties; each county is represented by a dot in the graph (reproduced by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc., from Forman *et al.*, 1990).



In [Example 5.12](#), information on the average prevalence of *H. pylori* IgG antibodies ('exposure') and on gastric cancer mortality ('outcome') for samples of individuals in each of 46 rural counties (the groups of interest) was obtained from surveys previously conducted in China.

The main advantage of ecological studies is that they can be done quickly and inexpensively. For instance, the question 'does a high-fat diet increase the risk of breast cancer' can be addressed in a much quicker and cheaper way in an ecological study than using any other study design ([Examples 5.2](#), [5.5](#) and [5.7](#)).

Although ecological studies are useful for generating hypotheses, they are of limited value in assessing whether there is a true exposure–outcome relationship at an individual level, since their results refer to *groups* of people and cannot be extrapolated to individuals. Attempts to infer associations at the individual level from those observed at the group

level are subject to the ‘*ecological fallacy*’ (see Section 11.2). For instance, in [Example 5.11](#), we do not know whether the women who developed breast cancer in each country were those who actually consumed a high-fat diet.

Despite these limitations, ecological studies may be the best approach to study exposures that are easier to measure at a group rather than at an individual level, such as air pollution and water quality. They are also useful for monitoring the effectiveness of population interventions such as health education campaigns and mass screening programmes (see Chapter 16).

5.2 Measures of exposure effect

The basic aim of an epidemiological study is to quantify the association between the exposure and the outcome of interest. To achieve this, the incidence of disease in a group of individuals exposed to the putative risk factor must be compared with the incidence in a group of persons not exposed. This comparison can be summarized by calculating either the ratio of the measures of disease occurrence for the two groups, which indicates the likelihood of developing the disease in the exposed individuals *relative* to those unexposed, or the difference between the two, which provides information about the *absolute* effect of the exposure in those exposed compared with those unexposed.

5.2.1 Relative measures of exposure effect

These measures estimate the magnitude of the association between exposure and disease, indicating how much more likely the exposed group is to develop the disease than the unexposed group. Three types of relative measure can be calculated:

$$\text{Risk ratio} = \frac{\text{Risk in the exposed group}}{\text{Risk in the unexposed group}}$$

$$\text{Rate ratio} = \frac{\text{Incidence rate in the exposed group}}{\text{Incidence rate in the unexposed group}}$$

$$\text{Odds ratio (of disease)} = \frac{\text{Odds of disease in the exposed group}}{\text{Odds of disease in the unexposed group}}$$

These measures are often collectively called measures of *relative risk*.

The *relative risk* is used as a measure of *etiological strength*. A value of 1.0 indicates that the incidence of disease in the exposed and unexposed

Example 5.13. Suppose that 20 000 workers were recruited into a cohort study. At the time of their entry into the study, individuals were classified as exposed or unexposed to a particular chemical substance on the basis of the type of job they had at that time. The whole cohort was then followed up for a period of two years to establish whether those exposed had an increased risk of dying from cancer. The results are presented in Table 5.1.

Table 5.1.

Hypothetical data from a cohort of 20 000 individuals followed up for a period of two years.

	Exposure	
	Yes	No
No. initially at risk	4000	16 000
Deaths	30	60
Person-years at risk ^a	7970	31 940
Risk ratio = $\frac{30/4000}{60/16\ 000} = \frac{7.5\ \text{per}\ 1000}{3.75\ \text{per}\ 1000} = 2.0000$		
Rate ratio = $\frac{30/7970\ \text{pyrs}}{60/31\ 940\ \text{pyrs}} = \frac{3.76\ \text{per}\ 1000\ \text{pyrs}}{1.88\ \text{per}\ 1000\ \text{pyrs}} = 2.0038$		
Disease odds ratio = $\frac{30/(4000-30)}{60/(16\ 000-60)} = \frac{0.00756}{0.00376} = 2.0076$		

^a Assuming that on average all deaths occurred in the middle of the follow-up period.

groups is identical and thus that there is no observed association between the exposure and the disease. A value greater than 1.0 indicates a positive association, or an increased risk among those exposed. Similarly, a relative risk less than 1.0 means that there is an observed inverse association or a decreased risk among those exposed. In [Example 5.13](#), workers exposed to the particular occupational hazard were twice as likely to die from cancer as those who were not exposed.

Note that in the above example the three measures of effect give similar estimates of relative risk. Death from cancer is a rare occurrence and therefore the number at risk remains practically constant throughout the study, since the cases represent a negligible fraction of the population. In practice, the three measures of effect will yield similar estimates of relative risk only for rare conditions (e.g., cancer). The estimates obtained by these three measures may differ considerably when a common disease (e.g., most infectious diseases) is examined or when a moderately rare disease is studied over a long period of time (e.g., coronary heart disease in women followed over 20 years).

5.2.2 Absolute measures of exposure effect

Information on the relative risk alone does not provide the full picture of the association between exposure and disease. Table 5.2 shows relative risks, calculated as rate ratios, of diseases A and B among those exposed to a certain risk factor. Although the rate ratio is higher for disease A, the incidence rate for disease A is increased by only 15 cases per 100 000 person-years at risk, whereas the incidence rate for disease B is increased by 40 cases per 100 000 person-years at risk. Clearly, the absolute impact of the exposure, measured by the rate difference, is quite different for these two diseases.

The *excess risk* (also called *attributable risk*) is an absolute measure of exposure effect. It indicates how many extra cases the exposure is respon-

	Disease A	Disease B
Incidence rate in the exposed group ^a	20	80
Incidence rate in the unexposed group ^a	5	40
Rate ratio	4.0	2.0
Rate difference ^a	15	40

^a Rates per 100 000 pyrs.

Table 5.2.

Incidence of disease A and B among persons exposed and persons unexposed to a particular risk factor, rate ratios and rate differences: hypothetical data.

sible for, *assuming that the relationship between exposure and disease is causal*. It can be calculated either as the difference of risk (*risk difference*) or of rates (*rate difference*).

Excess risk is equal to:

$$\text{Risk difference} = \text{risk in the exposed} - \text{risk in the unexposed}$$

or

$$\text{Rate difference} = \text{rate in the exposed} - \text{rate in the unexposed}$$

The excess risk is especially useful in evaluating the impact of introduction or removal of a risk factor and its value can be translated into the number of cases of the disease among the exposed that could be prevented if the exposure were completely eliminated. In the above example, the importance of the exposure as an etiological factor was given by the rate ratio and it was greater for disease A than for disease B, but, from a public health viewpoint, the exposure is much more important for disease B because more cases of disease B than of disease A would be avoided if the exposure were removed (assuming that the public health costs of a case of disease A are similar to those of disease B). Thus, relative risk measures the *strength* of the association between the exposure and the outcome of interest, whereas the excess risk measures the *impact* of the association in public health terms. In

contrast to the relative risk, however, the magnitude of the excess risk cannot be generalized to other populations because it depends on the baseline incidence in the unexposed group, which tends to vary between populations.

It is useful to express the excess risk in relation to the risk (or rate) in the exposed group. This measure is called the *excess fraction* (also known as the *excess risk percentage* or *attributable risk percentage*) and can be calculated as

$$\text{Excess fraction (\%)} = 100 \times (\text{excess risk/risk (or rate) in the exposed})$$

In the above example, the excess fraction for diseases A and B would be

$$\text{Excess fraction (\%)} \text{ for disease A} = 100 \times (15 \text{ per } 100\,000 \text{ pyrs} / 20 \text{ per } 100\,000 \text{ pyrs}) = 75\%$$

$$\text{Excess fraction (\%)} \text{ for disease B} = 100 \times (40 \text{ per } 100\,000 \text{ pyrs} / 80 \text{ per } 100\,000 \text{ pyrs}) = 50\%$$

This excess fraction represents the proportion of cases among the exposed that can be attributed to the exposure (assuming causality). In other words, it represents the proportion of cases among the exposed that could have been prevented if they had never been exposed. Thus, 75% of the cases of disease A and 50% of the cases of disease B could have been prevented among the exposed if they had never been exposed.

Example 5.14. A total of 34 439 British male doctors were followed up for 40 years and their mortality in relation to smoking habits was assessed (Doll et al., 1994a). Mortality from certain diseases is shown in Table 5.3.

Table 5.3.

Mortality by smoking habits from selected causes among male British doctors.^a

Underlying cause of death	Never smoked regularly Rate ^b (1)	Current cigarette smoker Rate ^b (2)	Rate ratio (2)/(1)	Rate difference ^b (2)–(1)	Excess fraction (%) (2)–(1) x 100 (2)
Cancer					
All sites	305	656	2.2	351	54
Lung	14	209	14.9	195	93
Oesophagus	4	30	7.5	26	87
Bladder	13	30	2.3	17	57
Respiratory diseases (except cancer)	107	313	2.9	206	66
Vascular diseases	1037	1643	1.6	606	37
All causes	1706	3038	1.8	1332	44

^a Data from Doll et al., 1994a.

^b Age-adjusted rates per 100 000 pyrs.

Example 5.14 shows that 44% of the deaths that occurred among male British doctors who smoked could be attributed to smoking (assuming causality). The proportion of deaths that could be attributed to smoking varied by disease. For those diseases shown in **Table 5.3**, this proportion was highest for lung cancer (93%) and lowest for vascular diseases (37%). However, if smokers had never smoked, the total number of deaths prevented would have been much greater for vascular diseases (606 per 100 000 pyrs) than for lung cancer (195 per 100 000 pyrs).

Similar absolute measures of effect can be calculated when those exposed have a lower risk of developing the disease than those unexposed. In these circumstances, we would have

Risk reduction = risk (or rate) in the unexposed – risk (or rate) in the exposed

Prevented fraction (%) = $100 \times (\text{risk reduction} / \text{risk (or rate) in the unexposed})$

Example 5.15. Suppose that a group of oral contraceptive users and a group of never users were followed up in time and their ovarian cancer incidence was measured and compared. The results from this hypothetical study are shown in **Table 5.4**.

	Oral contraceptive use	
	Ever	Never
Ovarian cancer cases	29	45
Person-years at risk	345 000	321 429
Rate per 100 000 pyrs	8.4	14.0
Rate ratio = 8.4 per 100 000 pyrs / 14.0 per 100 000 pyrs = 0.60		
Risk reduction = 14.0 per 100 000 pyrs – 8.4 per 100 000 pyrs = 5.6 per 100 000 pyrs.		
Prevented fraction (%) = $100 \times (5.6 \text{ per } 100\,000 \text{ pyrs} / 14.0 \text{ per } 100\,000 \text{ pyrs}) = 40\%$.		

Table 5.4.

Distribution of ovarian cancer cases and person-years at risk by oral contraceptive use: hypothetical data.

In **Example 5.15**, 40% of ovarian cancer cases could have been prevented among never-users if they had used oral contraceptives.

5.3 Conclusions

Before implementing a study, careful consideration must be given to the appropriateness of the proposed study design, especially in terms of practical feasibility, information to be obtained, expected duration of the study and total costs. The advantages and disadvantages of each of these study designs are covered in more detail in subsequent chapters (7–11).

Further reading

* References to books and papers dealing with each study design are given in the relevant chapters.

All the measures of effect discussed in this chapter can be directly calculated from intervention and cohort studies, since the incidence of disease in those exposed and in those unexposed is known. This is not the case in other study designs, where these measures of effect can be estimated only indirectly. For instance, in case-control studies, the subjects are selected on the basis of their disease status (sample of subjects with ('cases') and without ('controls') a particular disease), not on the basis of their exposure status. Therefore, it is not possible to calculate the incidence in exposed and unexposed individuals. It is, however, possible to calculate the odds of exposure among cases and the odds of exposure among controls and obtain an odds ratio of exposure (see Chapter 9). It can be shown that, depending on the sampling scheme used to select controls, this measure provides an unbiased estimate of one of the three relative measures of effect considered in Section 5.2.1. This is discussed in detail in later chapters (9-11, 16).

Box 5.1. Key issues

- There are two main types of epidemiological study: intervention (experimental) and observational.
- In *intervention studies*, the allocation of the study subjects to the different study groups is implemented by the investigator. Thus, if conducted properly, the intervention and the control groups would be similar in all respects apart from the exposure under study. There are two types of intervention study:

Clinical trials, where the main aim is to assess the value of new forms of treatment.

Field trials, where the objective is to evaluate whether an intervention decreases the risk of disease among disease-free people. Field trials can be conducted at an individual level, if the unit of allocation is the individual, or at an aggregated level, if the unit is a group of people. Community trials are an example of trials carried out at an aggregated level, where whole communities are the unit of allocation.

- In *observational studies*, the researchers limit themselves to observing the occurrence of disease in people who are already segregated into different exposure groups. There are various types of observational study:

Cohort studies, in which a study population (or 'cohort') is selected and the exposure status of its members assessed. The cohort is followed up in time and the occurrence of disease in the different exposure groups is measured and compared.

Case-control studies, in which a group of patients with a particular disease or condition ('cases') and a suitable group of subjects without that disease or condition ('controls') are selected and their past exposure to putative risk factors is ascertained and compared

Box 5.1. Key issues (Contd)

Cross-sectional surveys, in which a sample of subjects from a defined population is selected and information on the exposure(s) and outcome(s) of interest is collected simultaneously at a single point in time.

Routine-data-based studies, in which the data are derived from routine data-collection systems (e.g., cancer registration or death certification). They may be carried out at an individual level if information on the exposure(s) and outcome(s) of interest is available for each of the study subjects or at an aggregated level (ecological studies) if the group rather than the individual is the unit of study.

- Once the data from a particular study have been collected, the association between the exposure and the outcome of interest can be quantified by calculating an appropriate *measure of effect*. This may be expressed as either the ratio of the measure of disease occurrence in the exposed relative to that in the unexposed (*relative measure*) or as the difference between the two (*absolute measure*). The first type of measure is particularly important when assessing etiology, whereas the second type is more useful for evaluations of the public health impact of the association.
- Each study design has its own limitations and strengths. These are considered in detail in subsequent chapters of this book (Chapters 7–13).