Overview

Epidemiology is the cornerstone of public health. It employs rigorous methods and a quantitative approach to study the health of populations rather than individuals. Epidemiological methods are used to identify the causes of poor health, measure the strength of association between causes and outcomes, evaluate interventions and monitor changes in population health over time. The study of epidemiology provides the evidence-base for appropriate public health policy, planning and practice. This chapter provides an introduction to the key approaches of epidemiological research.

Learning objectives

When you have completed this chapter you should be able to:

- describe the key aspects of epidemiology
- discuss the complex factors involved in the study of causality
- identify the basic study designs used in epidemiology
- recognize the role of epidemiology in society.

What is epidemiology?

Epidemiology is the study of the distribution and determinants of health states or events in specified populations, and the application of this study to control health problems (adapted from Porta and International Epidemiological Association, 2008). Health states or events usually refer to infection, illness, disability, or death but may equally be used to refer to a positive outcome (e.g. survival). Epidemiological studies describe the distribution of these health outcomes in terms of frequency and pattern. The frequency is the number of occurrences of an outcome within a given time period, and the pattern refers to the occurrence of the outcome by time, place and personal or population characteristics. Determinants influence the frequency and pattern of health outcomes and are known as risk factors or protective factors, depending on whether they result in a negative or positive health outcome respectively.

Epidemiological research also involves the testing of preventive interventions (e.g. vaccines, improved hygiene) and therapeutic interventions (e.g. medicines, surgery) to improve health and survival. An intervention may be evaluated either under ideal (research-controlled) conditions to assess its efficacy or through a routine delivery system to assess its effectiveness.
After collecting epidemiological evidence, its application to improve health is a natural progression. Identification of risk factors and protective interventions, and quantification of their effects are key to informing action. Knowledge of the distribution and time-trends of outcomes, risk factors, and intervention coverage may be used for advocacy, for health promotion, and to inform public health policy and practice.

The study of epidemiology

The two main approaches to epidemiological study are descriptive and analytical (Figure 1.1). Descriptive epidemiology may provide information on the distribution of health outcomes by age, population type, geography or over time. Sources of descriptive data include routine monitoring such as registers of births and deaths, notification systems of specific diseases or adverse treatment reactions, and hospital or clinic records. Population censuses may also provide data on births, deaths, and a variety of risk factors (e.g. age, gender), and there is an overlap with demography (i.e. research on changes in the size, structure and distribution of human populations). Population health surveys evolved from censuses and provide information on the use of health services, coverage of interventions and the frequency of specific outcomes.

Figure 1.1 Main sources of epidemiological data
Source: Ilona Carneiro.

Activity 1.1

Figure 1.2 shows a declining trend in the incidence of rheumatic fever in Denmark since 1900. Briefly describe what this might suggest.
Analytical epidemiology aims to investigate which factors may be responsible for increasing or decreasing the probability ('risk') of an outcome. Identifying the cause of an outcome is not always simple, and can be described in terms of **sufficient cause** and **component causes**. Sufficient cause refers to a factor or set of factors that inevitably produces the outcome. The factors that form a sufficient cause are called component causes. Some component causes are essential for the outcome to occur: tuberculosis cannot occur without *Mycobacterium tuberculosis*, and this is known as a **necessary cause**. However, some people may be infected with *M. tuberculosis* without developing tuberculosis, because other components such as immune status and concurrent infections (e.g. HIV) will determine their susceptibility to the disease.

A single necessary cause is rarely sufficient to cause the outcome. While this may make epidemiological investigation of causality more difficult to untangle, it works to our advantage in public health, as it means that there are often several points at which we can intervene to reduce the likelihood of an outcome. Necessary causes may be:

1. infectious agents such as viruses, bacteria or parasites;
2. environmental agents such as sun-rays or allergens (e.g. pollen, dust-mites);
3. industrial agents such as chemicals (e.g. nicotine) or radiation (e.g. mobile phones);
4. genetic factors such as chromosomal abnormalities;
5. physical factors such as violence or car accidents;
6. psychological factors such as stress or abuse.
Key principles of epidemiology

Component causes may influence an individual's contact or response to a necessary cause. Environmental factors tend to affect contact and may be physical (e.g. climate, altitude), biological (e.g. vectors that transmit an agent) or structural (e.g. crowding, sanitation). Human factors affect both contact and response, and include age, sex, ethnicity, behaviour, genetics, and nutritional and immunological status. These environmental and human factors also interact, making the whole process even more complex. For example, people living in conditions of poor sanitation will have greater contact with the polio virus because transmission is mainly via faecal contamination. Children will be at greater risk of infection than adults because of their poorer sanitary practices and also because of their lack of natural immunity or incomplete immunization.

As you may have realized, depending on the perspective we take, a cause can also be considered as an outcome for the purpose of epidemiological investigation. For example, human immunodeficiency virus (HIV) is a necessary cause of acquired immunodeficiency syndrome (AIDS). However, we might then want to consider HIV infection as an outcome, and identify the necessary cause as unprotected sex with an infected individual, or contact with contaminated needles. This leads us to consider other risk factors that might increase the likelihood of HIV infection: multiple sexual partners, sharing of intravenous drug needles or poor safety practices in health facilities. However, while these risk factors can be component causes, they are not necessarily causal. A person may become infected through only one sexual contact, while another person with multiple sexual partners may not become infected at all.

Relating a causative agent or risk factor – from here on termed exposure – to an outcome of interest is known as inferring causality. For an association to be causal, the exposure must occur before the outcome. Other factors that support a causal relationship include a dose–response relationship, the strength of the association seen, a plausible biological mechanism of action, and reproducibility of the result. These and other factors that support causality will be discussed in more detail in Chapter 4.

**Alternative explanations**

Analytical methods may confirm an association between an exposure and outcome, but causality can only be inferred if alternative explanations, namely chance, bias and confounding, have been accounted for. Chance is the possibility that there is random error and is usually reduced by increasing sample size, using random selection (observational studies) or randomization (intervention studies). Bias refers to systematic differences between comparison groups, which may misrepresent the association being investigated. Confounding is caused when another factor, independently associated with both the outcome and exposure of interest, influences the association being investigated. These alternative explanations for an apparent association between an outcome and exposure, and the challenges of inferring causality, will be discussed in Chapter 4.

**Analytical study designs**

The main approach to investigating causal relationships is through analytical studies. An epidemiological investigation starts with the development of a hypothesis, which takes the form of a proposed association that can be tested. For example, 'smokers are at a higher risk of lung cancer'. An analytical study will then aim to find out whether there is sufficient or insufficient evidence to support this hypothesis.
Analytical epidemiology takes two forms: observation or intervention (Figure 1.1). An **observational study** aims to compare the frequency of the outcome in groups or individuals with and without the exposure of interest. An intervention study is effectively an experiment, and therefore restricted to evaluating the effect of reducing a risk factor or increasing a protective factor on the frequency of an outcome. Five study designs form the core of epidemiological research and these will be considered in more detail in Chapters 6–10:

1. **Ecological studies** consider populations of individuals and aim to relate the total frequency of an outcome to an average level of exposure by population group. For example, differences in alcohol consumption and incidence of breast cancer by country.
2. **Cross-sectional studies** collect data on outcome and exposure at one point in time from a random sample of study **subjects**. For example, prevalence of HIV in relation to male circumcision.
3. **Cohort studies** compare individuals with recorded differences in exposure to measure the occurrence over time of the outcome in relation to exposure. For example, incidence of cervical cancer in women with and without human papillomavirus infection.
4. **Case-control studies** identify individuals with and without the outcome, and examine whether they differ in relation to previous exposure. For example, mobile telephone use among people with brain tumours compared to those without brain tumours.
5. **Intervention studies** allocate a protective factor to individuals or groups, and compare the frequency of the outcome in those exposed with those unexposed. For example, the incidence of malaria among children given an insecticide-treated mosquito net, compared with those given an untreated mosquito net. Intervention studies may be randomized or non-randomized.

**The application of epidemiology**

The data and relationships identified through epidemiological study may be used in various ways. Descriptive epidemiological methods enable health professionals to identify the actual and potential health problems in a population. The burden of health outcomes or associated risk factors can be quantified, related to existing health services, and tracked to predict changes over time. An overview of the health issues affecting a population, and more importantly the relative distribution of these outcomes, enables priorities to be set and programmes to be planned. For example, Figure 1.3 shows that the incidence of road traffic deaths among children was estimated to be greater in Africa than in Europe for 2004, highlighting the need for relevant interventions in Africa.

Once a risk factor has been identified through analytical studies, health promotion activities may be developed to reduce exposure to the outcome at the individual level (e.g. encouraging smokers to stop smoking through education and support programmes) or population level (e.g. banning smoking in public places). Screening programmes may be implemented to increase early diagnosis and appropriate treatment (e.g. recommended mammograms for all women over 50 years old, who are at greater risk of breast cancer than younger women).

Monitoring and evaluation of health programmes are necessary to assess whether an implemented intervention is safe and effective under routine conditions, and whether
Activities 1.2–1.5

To help you put some of these epidemiological ideas into context, you will now look at a famous example from the nineteenth century.

John Snow (1813–58), a distinguished physician and considered one of the fathers of epidemiology, is best known for his studies of cholera outbreaks in London between 1848–49 and in 1854 (Snow 1936). These are the first documented epidemiological investigations. Activities 1.2–1.5 use John Snow’s cholera studies to illustrate the epidemiological approach, from descriptive epidemiology and hypothesis generation (1.2), to hypothesis testing and refinement (1.3–1.4), and application of epidemiological data (1.5). After attempting each activity you should refer to the feedback at the end of the chapter to prepare you for the next activity. Feedback does not provide the only true answers, as there are many accurate ways to answer these questions.

Cholera periodically swept across Europe during the nineteenth century. Cholera was characterized by profuse painless diarrhoea and clear fluid vomit that caused rapid dehydration, but the cause was unknown. After a severe epidemic in London in 1832, cholera reappeared in 1848. The first definite case was a seaman, newly arrived from Hamburg where cholera was prevalent. He died a few hours after the onset of symptoms on 22 September 1848 in a hotel near the River Thames. The next case was a lodger in the same room, who developed cholera symptoms on 30 September 1848. During the epidemic, approximately 15,000 deaths were recorded. Cholera mortality
Principles of epidemiology

in this epidemic was particularly high in residential areas downstream from the hotel, and decreased progressively upstream.

Microorganisms had not yet been discovered and one of the popular beliefs about disease causation was the 'miasma' theory – that breathing bad air caused disease. John Snow had previously documented several instances in which people had come into contact with cases of cholera and developed the disease within a few days. While investigating several case series of cholera, he made the following observations:

- Cholera was more readily transmitted within poor households and to those who had handled a case of cholera.
- Miners had suffered more than any other occupation.
- Almost no doctor who attended to cholera cases or conducted post-mortems had developed cholera.
- Most cases of cholera developed within 24–48 hours after contact with a case of cholera.
- Cholera was characterized by profuse painless diarrhoea and often proceeded with so little feeling of general illness that patients did not consider themselves in danger, or seek advice, until the illness was far advanced.

Activity 1.2

1. If you were a doctor in Snow's time, list what hypotheses you might generate about cholera transmission from these observations.
2. Describe how Snow might have interpreted his observations to oppose the 'miasma theory' and support alternative hypotheses such as those in question 1.
3. Identify what the most plausible explanations are for the observed association between elevation of residential area and level of mortality from cholera.

Activity 1.3

During the nineteenth century, private companies that obtained water directly from the river Thames supplied the drinking water in London. Each company had its own network of pipes. In some areas these networks overlapped and different companies could supply houses along a single street. The Southwark and Vauxhall (S&V) Company and the Lambeth Company were the two major water suppliers to the cholera-affected areas during the epidemics that John Snow investigated.

Between 1849 and 1853, when London was free of cholera, the Lambeth Company moved its water source upstream to an area outside London, while the S&V Company continued to draw water from a downstream source in London. Snow collected data on the number of houses supplied by the S&V Company and the Lambeth Company. When the cholera epidemic recurred in London in 1854, he collected data on sources of water for households of those who died of cholera. Table 1.1 shows the number of cholera deaths per 10,000 households, stratified by water source, during the first seven weeks of the epidemic.

1. Do the data presented in Table 1.1 support Snow's hypothesis that cholera is transmitted through water? Give reasons for your answer.
Table 1.1 Water sources and cholera mortality in London, 9 July to 26 August 1854

<table>
<thead>
<tr>
<th>Source of water</th>
<th>Total number of households</th>
<th>Number of cholera deaths</th>
<th>Deaths per 10,000 houses</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;V Company</td>
<td>40,046</td>
<td>1,263</td>
<td>315</td>
</tr>
<tr>
<td>Lambeth Company</td>
<td>26,107</td>
<td>98</td>
<td>37</td>
</tr>
<tr>
<td>Rest of London</td>
<td>256,423</td>
<td>1,422</td>
<td>59</td>
</tr>
</tbody>
</table>

Source: Adapted from John Snow (1936).

2 Are these data adequate to conclude that cholera mortality is higher in houses supplied by the S&V Company than in houses supplied by the Lambeth Company? Give reasons for your answer.

3 What further questions might you ask before reaching any conclusions based on these data?

Activity 1.4

Snow investigated a severe outbreak of cholera in the Soho area of London. He collected house addresses of all 616 recorded cholera deaths between 19 August and 30 September 1854. From these data, he produced a map showing the distribution of cholera deaths (bars) and positions of the water pumps (filled-circles) used by Soho households (Figure 1.4). This is known as a ‘spot map’.

Figure 1.4 Distribution of cholera deaths around Golden Square, London, August–September 1854 presented in a ‘spot map’.

Source: John Snow (1936).
1 Describe the distribution of cholera deaths in relation to the position of water pumps in Figure 1.4.
2 What might explain the differences in the distribution of deaths around water pumps A, B, C and D?
3 Can you conclude that water from a particular pump was the source of the cholera epidemic? Give reasons for your answer.
4 What further information do you need?

Activity 1.5

Snow discovered that a brewery was located in the two blocks with no cholera deaths with a deep well on the premises. The brewery workers and people living nearby collected water from the brewery well. Additionally, brewery workers had a daily quota of malt liquor. This information convinced Snow that pump B was the source of the cholera. He persuaded the local authorities to remove the pump-handle, preventing further use of the pump after the 8th September.

The dates of onset of symptoms of the 616 fatal cases of cholera recorded between 19 August and 30 September are shown in Figure 1.5.

1 Describe what the graph in Figure 1.5 shows.
2 Did removal of pump B end the cholera epidemic? Explain your answer.
3 What other factors might explain why the epidemic stopped?

Figure 1.5 Distribution of cases of cholera by date of onset (based on data from Snow, 1936).
Source: Drawn by Ilona Carneiro using data from John Snow (1936).
Conclusion

Epidemiology includes both a scientific approach (evidenced-based medicine) and a societal perspective (population-based studies and solutions) to health. You have been introduced to several new concepts and terms that are key to the understanding of epidemiology: descriptive epidemiology, inferring causality, analytical study designs, interpretation of results and the applications of epidemiology. These issues are fundamental for those involved in clinical and public health, and will be discussed further in subsequent chapters.

References

Snow J (1936) Snow on Cholera (being a reprint of two papers), London: Oxford University Press.

Feedback for activities

Activity 1.1

The graph in Figure 1.2 shows that the incidence of rheumatic fever drops particularly sharply after 1900, having been relatively steady for the previous 40 years. This suggests that some event might have triggered the decline in the incidence of rheumatic fever around 1900.

Rheumatic fever is caused by haemolytic streptococcal upper respiratory tract infection, which is associated with poverty and overcrowding. Therefore, it is reasonable to suggest that the decline in rheumatic fever was the result of improvements in socioeconomic conditions in Denmark that occurred at the beginning of the twentieth century.

Activity 1.2

1 You might have listed some or all of the hypotheses that Snow developed. Based on his observations, Snow generated the following hypotheses on the mode of transmission of cholera:
   • Cholera can be transmitted from the sick to the healthy.
   • Cholera is caused by some material (Snow called it ‘morbid matter’) that can increase and multiply in the body of the person it attacks.
   • The causative agent must be introduced into the alimentary canal by swallowing.
   • The causative agent may be transmitted through water from the sick to the healthy.

2 To dispute the miasma theory, Snow argued that:
   • The risk of transmission of cholera was high in miners and people of low socioeconomic status because these groups had poor hygiene practices and were, therefore, more likely to have contact with faecal matter from cholera patients than those of higher socioeconomic status.
• Few doctors developed cholera because they washed their hands after seeing each cholera patient.
• If transmission were through the air or via a vector, the disease would have been transmitted from cholera patients to more doctors.
• Thus, the disease was most likely caused by some causative agent transmitted by direct contact.

3 The observed association between higher elevation of residential area and lower mortality from cholera could support the theory of bad air causing cholera. However, as Snow argued, the water downstream of the initial cases was more likely to be polluted with sewage than the water upstream. Thus, the increased risk of cholera transmission in the areas downstream also supported his theory that the causative agent was most probably transmitted through water.

Activity 1.3

1 Yes, these data support Snow’s hypothesis that cholera is transmitted through water, but they do not prove it. For example, no information on other possible modes of transmission is included.

2 The risk of cholera death was 315/10,000 in houses supplied by the S&V Company, 38/10,000 in households supplied by Lambeth Company, and 55/10,000 in houses supplied by other sources. These data suggest that the risk of cholera death was 8–9 times higher (315/38) in households supplied by the S&V Company than in households supplied by the Lambeth Company. However, they do not prove causality and are not adequate to make firm conclusions that cholera mortality is higher in households supplied by the S&V Company.

3 Before reaching any conclusions, you would want to consider whether the number of people per household, their socioeconomic status, and other potential factors associated with the risk of transmission of cholera are comparable between these two populations. For example, the S&V Company might have supplied water to multiple-occupancy buildings while Lambeth supplied individual family houses. If this were the case, then the risk of cholera death per house between the two populations would not be comparable since the average number of people per house would differ between them. Since the S&V Company was drawing water from downstream, it is possible that households supplied by the company would have been in downstream areas and might be poorer than households upstream. Thus, although these data appear to support Snow’s hypothesis, more information is needed to be convincing.

Activity 1.4

1 Figure 1.4 shows that there was a cluster (collection) of many deaths around pump B (the Broad Street pump), very few deaths near pumps A and D, and almost no deaths around pump C.

2 If water from all the pumps was the source of cholera, there would probably have been similar numbers of deaths around each pump rather than more around pump B. However, it is possible that people did not drink the water from pumps A, C and D due to bad taste, smell, or inconvenience, or that the water from these pumps might not have carried the cholera causative agent.

3 While pump B has the greatest spatial clustering of cholera deaths and might have been the source, this is not sufficient to conclude that pump B was the source of
cholera. Two blocks of buildings very close to pump B did not have a single death from cholera.

4 More information is needed to explain the absence of deaths in the two blocks nearby before implicating the water from pump B as the source of the epidemic. For example, if it could be shown that there was no death in these blocks for reasons such as:
   - No one lived there.
   - Inhabitants had alternative sources of water.
   - Inhabitants had some kind of protection against cholera.

Activity 1.5

1 There appears to have been a low background number of cases (zero or one case per day) before 30th August. There was an explosive rise in the number of cases over three days, which decreased to previous levels after 12 days. The most likely explanation for the sudden rise in the number of cholera deaths would be exposure of the population to a causal agent from a common source.

2 It is unlikely that removal of the pump-handle from pump B stopped the epidemic, because the number of cholera deaths had already dropped to almost the background level by the time the pump-handle was removed. However, removing the pump-handle B may have prevented another outbreak of cholera if pump B still contained the causal agent and was available for use.

3 There are several possible explanations for the end of the epidemic:
   - People who lived in the epidemic area might have moved elsewhere due to fear of contracting cholera.
   - All susceptible people (those who had no form of immunity to cholera and were therefore at-risk of infection) might have been exposed within a short time, leaving very few susceptible individuals.
   - The amount of causal agent in the water could have reduced.

*Vibrio cholera*, the bacterium that causes cholera, was identified by Robert Koch in 1883, several decades after Snow identified appropriate preventive measures from his epidemiological investigations:

> I feel confident, however, that by attending the above mentioned precautions (personal hygiene, boiling soiled bedclothes of patients, isolation and quarantine, improved waste disposal, drainage, provision of clear water), which I consider to be based on correct knowledge of the cause of cholera, this disease may be rendered extremely rare, if indeed it may not be altogether banished from civilized countries. (Snow 1936)
Measuring the frequency of outcomes

Overview

The occurrence of health outcomes (i.e. infection, illness, disability and death) will vary between populations, geographical areas and over time. Epidemiological studies quantify the frequency of health outcomes, which is the number of occurrences in a defined population over a defined time-period. In this chapter you are introduced to the epidemiological measures used to determine the frequency of outcomes: prevalence, risk (including attack rates), odds and incidence rates.

Learning objectives

When you have completed this chapter you should be able to:

• identify and define the four common frequency measures: prevalence, risk, odds and incidence rate
• calculate each of these frequency measures
• recognise the use of attack rates in the investigation of outbreaks
• estimate person-time at risk.

Defining a case

To measure the frequency of an outcome in a population, it is first necessary to have a clear definition or description to identify the outcome of interest. In some situations the outcome is obvious (e.g. all cause death), but often, standardized criteria are needed (e.g. severe anaemia may be defined as haemoglobin less than 5 grams per decilitre). Individuals with the outcome of interest are often referred to as ‘cases’. The criteria used to define them form the case definition, which may not be clinically defined. The outcome may refer to an event such as a car accident rather than an illness. A ‘case’ may occur only once per individual (e.g. death), more than once (e.g. pregnancy), or frequently (e.g. diarrhoeal disease). Epidemiologists count cases using clinical assessments, diagnostic tests, registry or clinic record entries, observation, or even self-reporting in population surveys.

Knowing the number of cases is not enough to allow any comparison or association to be made. If you were told that there were 75 cases of tuberculosis in village A and only 25 cases in village B, you might be tempted to conclude that tuberculosis was more common in village A than in village B. However, without knowing how many people live in each village, this comparison is impossible to make. Once cases have been defined and counted, it is necessary to count the number of individuals in the population from which cases were identified and the time-period in which the cases occurred, to calculate frequency.
Key principles of epidemiology

Measuring disease frequency

The most common measures of disease frequency (i.e. prevalence, risk, odds, incidence rate) vary according to how cases and time-period are considered. We will use the analogy of a children’s game of musical chairs, where everybody dances while the music is played, and chairs are removed at regular intervals when the music stops. Those left without a chair on which to sit have to stand to one side until the game finishes. We can explain the measures of frequency as follows, where the outcome is ‘sitting on a chair’:

1. **Prevalence** counts the number of children sitting at a specific time point (e.g. at 11:30a.m.), compared with the total participants in the game (standing and sitting) at that same time point.
2. **Risk** counts those sitting after a specified time-period (e.g. after 10 minutes), compared with the total participants at the start of the game.
3. **Odds** counts those sitting after a specified time-period (e.g. after 10 minutes), compared with those standing after the same time-period.
4. **Incidence rate** counts those sitting at any point during the game and the total time that each individual participates, allowing for children who join the game late or leave early.

---

### Prevalence

**Prevalence** is the number of existing cases in a defined population at a defined point in time divided by the total number of people in that population at the same point in time:

\[
\text{Prevalence} = \frac{\text{Number of cases at one time point}}{\text{Total number of individuals in the defined population at same time point}}
\]

Prevalence is a proportion and can never be greater than one. It is dimensionless, meaning that it has no units, so the term ‘prevalence rate’ is incorrect. Prevalence is usually presented as a percentage by multiplying the proportion by 100. Prevalence is sometimes referred to as **point prevalence** to distinguish it from **period prevalence**. Period prevalence refers to the number of existing cases identified during a specified, usually short, period divided by the total number of people in that population during the same period.

The prevalence of an outcome may be measured during population surveys or cross-sectional studies. Prevalence is useful to rapidly assess the frequency of an outcome in a community. For example, in a cross-sectional survey of 200 boys aged 5–10 years old in a low-income setting, 60 were found to be stunted, i.e. had a height-for-age lower than the average. The prevalence of stunting in this group would be calculated as \( \frac{60}{200} = 0.3 \), which would be presented as \( 0.3 \times 100 = 30\% \) of young boys in this population being stunted (an indicator of chronic malnutrition) at the time of the survey.

---

### Incidence

**Incidence** is the frequency of new (‘incident’) cases in a defined population during a specified time-period. Incidence may be measured in ecological or cohort studies. There are three different ways of considering incidence: risk, odds and incidence rate.
Measuring the frequency of outcomes

Risk

Risk is also known as cumulative incidence because it refers to the total number of new cases in a defined ‘population at risk’ over a specified period of time:

\[
\text{Risk} = \frac{\text{Number of new cases in a specified time-period}}{\text{Total number of individuals at risk in the population at the start of that time-period}}
\]

This measure can be interpreted as the likelihood (‘risk’) that an individual will develop an outcome during the specified time-period, and the ‘population at risk’ excludes existing (‘prevalent’) cases. Risk is also a dimensionless proportion, so can never be greater than one and has no units. However, its value can increase with the duration of the time-period under consideration, making it essential to specify the period at risk. For example, if a group of 100 people were studied for a year, and 75 had caught at least one cold during that year, we could say that the risk of catching a cold was \(75 \div 100 = 0.75\) or 75% in that year in that group. However, the result would be interpreted differently if 100 people had been studied for six months and 75 had caught at least one cold during this six-month period; it would have to be specified as a 75% risk over 6 months.

A specific form of risk used in disease outbreak settings is called the secondary attack rate. This is a misnomer, as it is a proportion and not a rate (see below), but the term is commonly accepted. The secondary attack rate is calculated as the number of new cases among contacts of a primary case in a specified period of time:

\[
\text{Secondary attack rate} = \frac{\text{Number of new cases among contacts in a specified time-period}}{\text{Total number of contacts of a primary case in that time-period}}
\]

This can be interpreted as the ‘risk’ that a contact of a case will develop the outcome during the specified time-period. The total number of contacts is often estimated from the household members of primary cases, but may also include school or workplace contacts. For example, if eight children developed varicella (chicken pox) in an outbreak at a school, and five out of a total of 15 siblings developed varicella in the subsequent two weeks, we could estimate the secondary attack rate, or risk of developing varicella among household contacts, as \(5 \div 15 = 0.33\) or 33% in this two-week time-period.

Odds

Odds is a different way of representing risk, and is calculated as the number of new cases divided by the number of individuals still at risk after a specified time-period:

\[
\text{Odds} = \frac{\text{Number of new cases in a specified time-period}}{\text{Number who did not become a case during that time-period}}
\]

The odds is actually a ratio of two proportions and can be greater than one. It is the ratio of the ‘risk’ that an individual develops the outcome during a specified time-period, to the ‘risk’ that the individual does not develop the outcome during that same time-period. Below you can see how this simplifies mathematically to the equation given above, as the denominator (total number at risk) is the same for both outcomes, and cancels-out:
Key principles of epidemiology

Odds = \frac{\text{Cases}}{\text{Total}} \div \frac{\text{Non-cases}}{\text{Total}} = \frac{\text{Cases}}{\text{Total}} \times \frac{\text{Total}}{\text{Non-cases}} = \frac{\text{Cases}}{\text{Non-cases}}

In the example, in which 75 people in a group of 100 caught a cold during a particular year, the odds of catching a cold would be calculated as $75 \div 25 = 3$. The odds of catching a cold would be 3 to 1, sometimes reported as 3:1, so that a person in that group would be three times more likely to catch a cold as not to catch a cold during that year.

Incidence rate and person-time at risk

Both risk and odds assume that the population at risk is followed over a specified time-period, and that all those who are included at the beginning of the time-period are counted at the end of the time-period. This is called a closed population. However, you might want to look at incidence in a dynamic or open population, in which people enter and exit the population at risk at different points and are therefore at risk for different lengths of time. Once the outcome has occurred, the individual will either no longer be at risk or, if the outcome can recur, there will be some interval of time before the individual is once more considered at risk. Therefore, instead of counting the total number of people at the start of the study, the time that each individual is at risk is calculated. This is known as the person-time at risk and is illustrated in Figure 2.1. People may start and stop being at risk at different times, due to births and deaths, immigration, acquiring the outcome, leaving the study population before the end (known as ‘lost to follow-up’), or reaching the end of the observation period.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.0</td>
</tr>
</tbody>
</table>

\* disease
\* lost to follow-up
\* time in study

Figure 2.1 Graphical representation of person-time at risk for 9 study participants during a 6-year study period

Source: Bailey et al. (2005).

The incidence rate allows us to account for variation in time at risk, and is calculated as the number of new cases divided by the total person-time at risk:
Measuring the frequency of outcomes

Incidence rate = \( \frac{\text{Number of new cases in a specified time-period}}{\text{Total person – time at risk during that time-period}} \)

This measure is a rate and is reported as the number of new cases per person-time at risk. It is essential to specify the time units, for example person-days, person-months, person-years, or more frequently 1,000 person-years at risk.

In Figure 2.1, the incidence rate is obtained by dividing the total number of cases by the total number of person-years at risk. Four people became cases during the study, so the incidence rate is \( 4 \div 40.5 = 0.099 \) cases per person-year at risk, or 99 cases per 1,000 person-years at risk (to avoid the use of too many decimal places).

In registry data or very large studies it might be difficult to know the exact person-time at risk for each individual in the population. In this situation, the population at the mid-point of the time-period of interest is multiplied by the period of time under consideration to give an estimate of the person-time at risk, so long as the population size does not change substantially over the study period. For example, data from a cancer registry (a record of all cancer cases notified by doctors) found 750 cases of breast cancer in a region of a politically-stable country between 2005 and 2010. The population of that region was recorded from a census to be 10,130 in 2007–2008. The person-time at risk is estimated as the mid-period population multiplied by the five-year period at risk. The incidence is calculated as \( 750 \div (10,130 \times 5) = 750 \div 50,650 = 0.015 \) per person-year at risk, or to reduce loss of precision due to rounding, \( 0.0148 \times 1,000 = 14.8 \) cases of breast cancer per 1,000 person-years at risk between 2005–2010.

Activity 2.1

Investigators were asked to determine the prevalence of malaria cases in two villages in rural Southeast Asia. There were 500 people (210 men and 290 women) in the two villages. Investigators spent a couple of days in November testing everyone in the villages for malaria, using rapid diagnostic tests (which detect malaria parasite-specific proteins in blood). Investigators found that 62 men and 22 women tested positive for malaria.

1. Write a simple case definition investigators could have used in this study.
2. Define the study population.
3. Based on your case definition and study population, what is the malaria prevalence among men and among women in the villages?

Activity 2.2

An outbreak of human monkeypox, a relatively rare orthopox viral disease similar to but milder than smallpox, was detected in a forested province of the Democratic Republic of Congo (DRC). Investigators found 5 secondary cases of monkeypox among 37 household contacts of a 9-year-old boy who was the first reported to have become infected.

1. Assuming all household contacts are equally at risk, what is the secondary attack rate among household contacts?
2. Data from previous outbreaks suggests that prior smallpox vaccination confers 85% protection from monkeypox. If all household contacts had been vaccinated against smallpox, how would this change your secondary attack rate calculation?
Investigation of an outbreak of measles in remote District X found 12 cases of measles had occurred among 350 children attending a local school over a one-month period. Each infected child came from a different household. The total number of additional children (i.e., household child contacts) in the 12 affected households was 67. Ten children in the affected households had previously had measles, while 20 children were reported to have received at least one dose of measles vaccine. However, immunization records were poor and many children had not been fully immunized. One month (i.e., approximately two incubation periods) later, four more children in the 12 affected households also developed measles.

1. What was the risk of measles in the school during the initial one-month period?
2. What is the secondary attack rate among household child contacts of the 12 children?
3. What does this tell us?

Investigators conducted a survey of intestinal worm infestation among 1,000 adolescent agricultural workers in Country Y. They found 620 adolescents were infested with one or more type of worm. After treating all adolescents found positive, investigators returned six months later and tested all 1,000 adolescents again. This time they found 390 adolescents infested.

1. Calculate the prevalence of worm infestation among adolescents in the first survey.
2. Calculate the risk and the odds of worm infestation among adolescents during the six-month study period.
3. What is the incidence rate of worm infestation among adolescents during the six-month study period?

One thousand men working in factory A were screened for HIV on 1 January 2010 and 50 of them were found to be HIV-positive. When the screening was repeated on the same 1,000 men on 1 January 2011, 62 men were positive, including the 50 men who were positive on the first screening. Nobody had died or been lost to follow-up.

1. What is the prevalence of HIV in men working in factory A on 1 January 2010, and on 1 January 2011?
2. What is the annual risk of developing HIV infection in men working in factory A during 2010?
3. What are the odds of developing HIV infection in men working in factory A during 2010?

One thousand men in factory B were screened for HIV on 1 January 2010 and 50 men were found to be HIV-positive. All of these men were tested for HIV at the end of each month until 31 December 2010. Twelve men became HIV-positive during this period,
while the remaining 938 men were still HIV-negative by 31 December 2010. Figure 2.2 shows when these 12 men became HIV-positive. Nobody died or was lost to follow-up during this period.

1 What is the total number of person-months at risk of HIV infection observed in this study?
2 What is the incidence rate of HIV infection among men working in factory B?
3 What are the odds of becoming infected with HIV in the first six months of 2010 compared with becoming infected in the last six months of 2010?

**Figure 2.2** Person-months at risk of HIV in factory B in 2010

Source: Bailey et al. (2005).

**Conclusion**

You have been introduced to the measures of prevalence and incidence (risk, odds and incidence rate) that are used to quantify the occurrence of an outcome in a defined population. These epidemiological measures of frequency help in assessing the public health importance of an outcome and planning appropriate health services. These measures also form the basis of analytical studies to investigate the association between exposures and outcomes.

**Reference**

Feedback for activities

Activity 2.1
1 You may have said that cases could have been defined as those individuals resident in one of the two villages who tested positive for malaria by rapid diagnostic test during the two days in November.
2 The study population includes all 500 individuals in the two study villages. There were 210 men and 290 women.
3 The number of prevalent cases among men is 62. There were 210 men tested. Therefore, the prevalence of malaria in men is $62/210 \times 100 = 0.295 \times 100 = 30\%$.
The number of prevalent cases among women is 22. There were 290 women tested. Therefore, the prevalence of malaria in women is $22/290 \times 100 = 0.076 \times 100 = 7.6\%$.

Activity 2.2
1 The number of cases of monkeypox among secondary contacts = 5. The total number of secondary household contacts = 37. Therefore, the secondary attack rate = $5/37 \times 100 = 0.135 \times 100 = 13.5\%$.
2 The total number of secondary household contacts estimated to be unprotected by vaccination = $37 - (37 \times 0.85) = 37 - 31.45 = 5.55$ (i.e. 6) unprotected household contacts. The number of cases of monkeypox among secondary contacts = 5.
   To calculate the secondary attack rate in this scenario we divide the number of cases among secondary contacts by the number of unprotected household contacts = $5/5.55 \times 100 = 0.90 \times 100 = 90\%$. This would indicate a dramatically more virulent monkeypox strain than did the first scenario.

Activity 2.3
1 The number of cases of measles among children at the school = 12. The total number of children at the school = 350. Therefore, the incidence risk = $12/350 \times 100 = 0.34 \times 100 = 3\%$.
2 The number of secondary cases of measles among household child contacts of the 12 children is 4. The total number of children at risk in the affected households = (total number of contacts – previous measles cases) = $67 - 10 = 57$. Therefore, the secondary attack rate among unvaccinated child contacts is the number of cases among household child contacts out of those child contacts who are actually known to be at risk (i.e. not already a case) = $4/57 \times 100 = 0.0702 \times 100 = 7\%$.
   In reality, there may be a higher number of contacts immune to measles through vaccination or prior disease.
3 An initial risk of (3%) and a secondary attack rate of (7%) indicate that the outbreak is increasing or that transmission is more efficient within households than within the school. As measles can be prevented by vaccination, the increase in cases suggests that vaccination coverage is insufficient and a mass measles vaccination campaign should be organized.

Activity 2.4
1 The number of prevalent cases in the first survey = 620. The total number of adolescents tested is 1,000. Therefore, the prevalence of infestation among adolescents in the first survey is $(620/1,000) \times 100 = 0.62 \times 100 = 62\%$.
2 The number of incident cases in the second survey = 390. The total number of adolescents tested is 1,000. Therefore, the risk of infestation among adolescents during the six-month period is \((390 \div 1,000) \times 100 = 0.39 \times 100 = 39\%\). The number of cases in the second survey = 390. The number of adolescents who are not cases = 1,000 – 390 = 610. Therefore, the odds of infestation among adolescents during the six-month period are \((390 \div 610) \times 100 = 0.64 \times 100 = 64\%\).

3 The number of incident cases in the second survey = 390. As we do not know how many months each adolescent contributed to person-time at risk, we need to use the mid-period population, calculated as the average of the population at start and end of the study (i.e. \((1,000 + 1,000)/2 = 1,000\), thus \(1,000 \times 6 = 6,000\) person-months), and assume that the adolescents did not leave the study area for significant periods during the six-month period. Therefore, the incidence rate of infestation among adolescents is estimated as \((390 \div 6,000) \times 1000 = 0.065 \times 1000 = 65\) per 1,000 person-months.

Activity 2.5

1 The number of prevalent cases at 1 January 2010 = 50. The total number of individuals tested in this population on 1 January 2010 = 1,000. Therefore, the prevalence at 1 January 2010 = \((50 \div 1,000) \times 100 = 5\%\).

The number of prevalent cases at 1 January 2011 is 62. The total number of individuals tested in this population at 1 January 2011 is 1000. Therefore, prevalence at 1 January 2011 = \((62 \div 1000) \times 100 = 6.2\%\).

2 The number of incident (new) cases in 2010 is 62 – 50 = 12. The number of individuals at risk of HIV infection on 1 January 2010 = 1000 – 50 = 950. Therefore, the annual risk of developing HIV infection in 2010 = \((12 \div 950 ) \times 100 = 1.26\%\).

The 50 men who were HIV-positive on 1 January 2010 are not included in the denominator because HIV-positive individuals do not become HIV-negative. Therefore those who were HIV-positive on 1 January 2010 were not at risk of developing HIV infection during 2010.

3 The number of incident cases during 2010 is 62 – 50 = 12. The number of individuals at risk of HIV infection in 2010 who did not become infected is 950 – 12 = 938. Therefore, the odds of developing HIV infection in 2010 = \(12 \div 938 \times 100 = 1.28\%\).

Activity 2.6

1 To calculate total person-months at risk, add total person-months contributed by men who remained HIV-negative and total person-months contributed while they were HIV-negative by men who subsequently became infected (as men who were already infected at the beginning of study do not contribute any person-months at risk).

Person-months at risk for men who remained HIV-negative throughout the study period = \(938 \times 12 = 11,256\). Person-months at risk for the 12 men infected during the study period = \(1 + (3 \times 4) + 5 + (2 \times 6) + 8 + 9 + (2 \times 11) + 12 = 81\). Therefore, total person-months at risk during the study period = \(11,256 + 81 = 11,337\).

Another way to look at this is that, as we do not know when exactly these men became infected, it would be more accurate to assume that 'on average' they contributed only 0.5 months at risk for the month prior to testing HIV-positive. Using this approach, total person-months at risk for the 12 men infected during the study period = \(0.5 + (3 \times 3.5) + 4.5 + (2 \times 5.5) + 7.5 + 8.5 + (2 \times 10.5) + 11.5 = 75\).
Using this approach, total person-months at risk during the study period would be $11,256 + 75 = 11,331$. In this particular example, as the overall person-time at risk is large, this will not make a difference to our estimate of incidence rate below.

2 To calculate an incidence rate, select an appropriate unit of person-time at risk for the denominator. In this example, person-months is an appropriate unit, as time at risk is already segmented into months and person-months contributed can be calculated easily from Figure 2.2.

Total number of incident cases = 12. Total person-months = 11,337. Therefore, incidence rate = $(12 ÷ 11,337) × 1,000 = 1.06$ per 1,000 person-months.

3 To calculate odds, divide the number of men who became HIV-positive in the first half of 2010 by the number who did not (i.e. who became positive in the second half of 2010).

Number of men who became HIV-positive in the first six months of 2010 = 7. Number of men who did not become HIV-positive during first six months = 5. Therefore, of those who became HIV-positive in 2010, the odds of becoming HIV-positive in the first 6 months = $7/5 = 1.4$. In other words, among those who became HIV-positive in 2010, the odds of becoming infected in the first half of the year were 40% higher (difference from an odds of 1, i.e. $1.4 – 1.0 = 0.4$ or 40%) than the odds of becoming infected in the second half of the year.