An Introduction to Veterinary Epidemiology *

Mark Stevenson †

EpiCentre, IVABS ‡

Massey University, Palmerston North, New Zealand

July 28, 2008

Contents

*Notes for Veterinary Biometrics and Epidemiology, taught as course 227.407 within the BVSc program at Massey University. Contributions from Dirk Pfeiffer, Cord Heuer, Nigel Perkins, and John Morton are gratefully acknowledged.

†M.Stevenson@massey.ac.nz

‡URL: http://epicentre.massey.ac.nz

1 Introduction

By the end of this unit you should be able to:

- Compare and contrast clinical and epidemiological approaches to disease management.
- Describe the factors that influence the presence of disease in individuals.
- Describe the factors that influence the presence of disease in populations.
- Explain what is meant by a common source and propagated epidemic (with examples). Explain how you would distinguish between a common source and propagated epidemic in a disease outbreak situation.

Epidemiology is the study of diseases in populations. Epidemiologists attempt to characterise those individuals in a population with high levels of disease and those with low levels. They then ask questions that help them discover what the high rate group is doing that the low rate group is not and *vice versa*. This allows factors influencing the risk of disease to be identified. Once identified, measures can be applied to reduce exposure to these risk factors — reducing the overall burden of disease in the population. This allows disease to be controlled even if the precise pathogenic mechanism (or the aetiologic agent) is not known.

It is useful to distinguish epidemiological from clinical approaches to disease management. The **clinical approach** is focussed on individual animals and is aimed at diagnosing a disease and then treating it. It involves physical examination and generation of a list of differential diagnoses. Further examinations, laboratory tests and possibly response to treatment are then used to narrow the list of differential diagnoses to a single diagnosis. In an ideal world this will always be the correct diagnosis. Research in health professionals has shown that the final diagnosis is nearly always drawn from the initial list of differential diagnoses. If the disease is not on the initial list of differentials then it tends not to become the final diagnosis. Diseases may be omitted from the list because the clinician is not familiar with them (exotic or unusual diseases) or because the disease is 'new' and has never been identified before.

The **epidemiological approach** to disease management is conceptually different in that there is no dependency on being able to precisely define the aetiological agent. It is based on observing differences and similarities between diseased and non-diseased animals in order to try and understand what factors may be increasing or reducing the risk of disease. In practice, clinicians unwittingly use a combination of clinical and epidemiological approaches in their day-to-day work. If the problem is relatively clear-cut then an epidemiological approach plays a very minor role. If the condition is new or more complex then the epidemiological approach is preferred since it will provide a better understanding of what makes individuals susceptible to disease and — once these factors are known — the measures required to control the disease become better defined.

1.1 Host, agent, and environment

Whether or not disease occurs in an **individual** depends on an interplay of three factors:

- The host;
- The agent; and
- The environment

The host is the animal (or human) that may contract a disease. Age, genetic makeup, level of exposure, and state of health all influence a host's susceptibility to developing disease. The agent is the factor that causes the disease (bacteria, virus, parasite, fungus, chemical poison, nutritional deficiency etc) — one or more agents may be involved. The environment includes surroundings and conditions either within the host or external to it, that cause or allow disease transmission to occur. The environment may weaken the host and increase its susceptibility to disease or provide conditions that favour the survival of the agent.

1.2 Individual, place, and time

The level of disease in a **population** depends on an interplay of three factors:

- Individual factors: what types of individuals tend to develop disease and who tends to be spared?
- Spatial factors: where is the disease especially common or rare, and what is different about these places?
- Temporal factors: how does disease frequency change over time, and what other factors are associated with these changes?

Individual

Individuals can be grouped or distinguished on a number of characteristics: age, sex, breed, coat colour and so on. An important component of epidemiological research is aimed at determining the influence of individual characteristics on the risk of disease. Figure 1 shows how mortality rate for drowning varied among children and young adults in the USA during 1999. The rate was highest in those aged 1 – 4 years: an age when children are mobile and curious about everything around them, even though they do not understand the hazards of deep water or how to survive if they fall in. What conclusions do we draw from this? Mortality as a result of drowning is highest in children aged 1 – 4 years: preventive measures should be targeted at this age group.

Place

The spatial pattern of disease is typically a consequence of environmental factors. Environmental factors include aspects of climate (temperature, humidity, rainfall) as well as aspects of animal management (management of animals in a certain area of a country may result in high rates of disease that may not be seen in other areas). Geographic Information Systems and easy access to spatial data (e.g. satellite images) have facilitated the ability to conduct spatial epidemiological analyses in recent years. Figure 2 shows the geographical distribution of BSE incidence risk in British cattle from July 1992 to June 1993. The amount and type of concentrate feeds fed to cattle is thought to have been responsible for the higher density of disease in the south of the country, compared with the north.

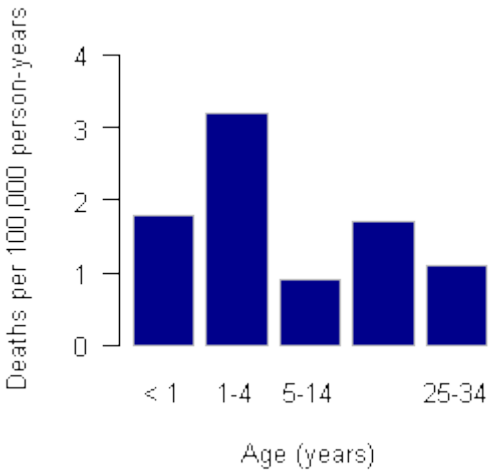


Figure 1: Mortality from drowning by age: USA, 1999. Reproduced from: Hoyert DL, Arias E, Smith BL, Murphy SL, Kochanek KD (2001) Deaths: final data for 1999. National Vital Statistics Reports volume 49, number 8. Hyattsville MD: National Center for Health Statistics.

Time

When talking about temporal factors influencing the pattern of disease we need to distinguish between animal referent time and calendar time. Animal referent time refers to the timing of events in relation to defined events that occur during an animal's lifetime. For example, we may talk of an increased risk of milk fever during the first 7 days of a lactation. Here, time is measured in relation to a calving event. Calendar time refers to the absolute timing of events. We may talk of the number of milk fever cases that occur in August, and compare those numbers with the number that occur in (say) December.

Temporal patterns of disease in populations are presented graphically using epidemic curves. An epidemic curve consists of a bar chart showing time on the horizontal axis and the number of new cases on the vertical axis, as shown in Figure 3. The shape of an epidemic curve can provide important information about the nature of the disease under investigation. An epidemic occurs when there is a rapid increase in the level of disease in a population. An epidemic is usually heralded by an exponential rise in the number of cases in time and a subsequent decline as susceptible animals are exhausted. Epidemics may arise from the introduction of a novel pathogen (or strain) to a previously unexposed (naïve) population or as a result of the re-growth of susceptible numbers some time after a previous epidemic due to the same infectious agent. Epidemics may be described as being either common source or propagated.

In a **common source epidemic**, subjects are exposed to a common noxious influence. If the group is exposed over a relatively short period then disease cases will emerge over one incubation period. This is classified as a common point source epidemic. The epidemic of leukaemia cases in Hiroshima following the atomic bomb blast would be a good example of a common point source epidemic. The shape of this curve rises rapidly and contains a definite peak at the top, followed by a gradual decline. Exposure can also occur over a longer period of time, either intermittently or continuously. This creates either an intermittent common source epidemic or a continuous common source epidemic. The shape of this curve rises rapidly (associated with

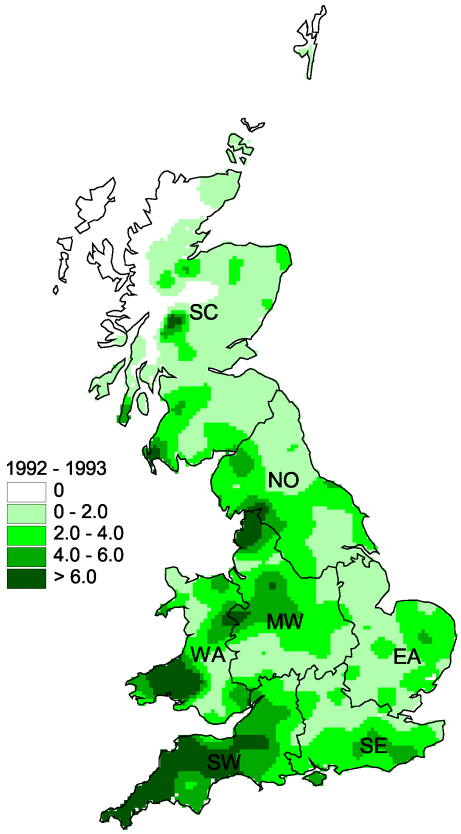


Figure 2: Incidence risk of BSE across Great Britain (expressed as confirmed BSE cases per 100 adult cattle per square kilometre), July 1992 – June 1993. Reproduced from Stevenson et al. (2000).

the introduction of the agent). The down slope of the curve may be very sharp if the common source is removed or gradual if the outbreak is allowed to exhaust itself.

A **propagated epidemic** occurs when a case of disease serves as a source of infection for subsequent cases and those subsequent cases, in turn, serve as sources for later cases. In theory, the epidemic curve of a propagated epidemic has a successive series of peaks reflecting increasing numbers of cases in each generation. The epidemic usually wanes after a few generations, either because the number of susceptibles falls below a critical level, or because intervention measures become effective.

Sometimes epidemic curves can show characteristics of being both common source and propagated. Figure 4 shows the epidemic curve for foot-and-mouth disease in the county of Cumbria (Great Britain) in 2001. This epidemic started as a common (point) source, then took on the characteristics of a propagative epidemic over time.

Endemic describes the situation when diseases (or events) occur at a predictable frequency. Figure 5 shows data from a descriptive study of dog and cat submissions to a humane shelter in Wellington, New Zealand from 1999 to 2006. In the plot on the left in Figure 5 there is a marked seasonal variation in the number of cats submitted to the shelter per month: no such pattern is apparent for dogs. If data are recorded over extended periods, long-term trends might be evident. In the plot on the right in Figure 5 it is evident that the number of dogs and cats submitted to the shelter decreased steadily throughout the study period.

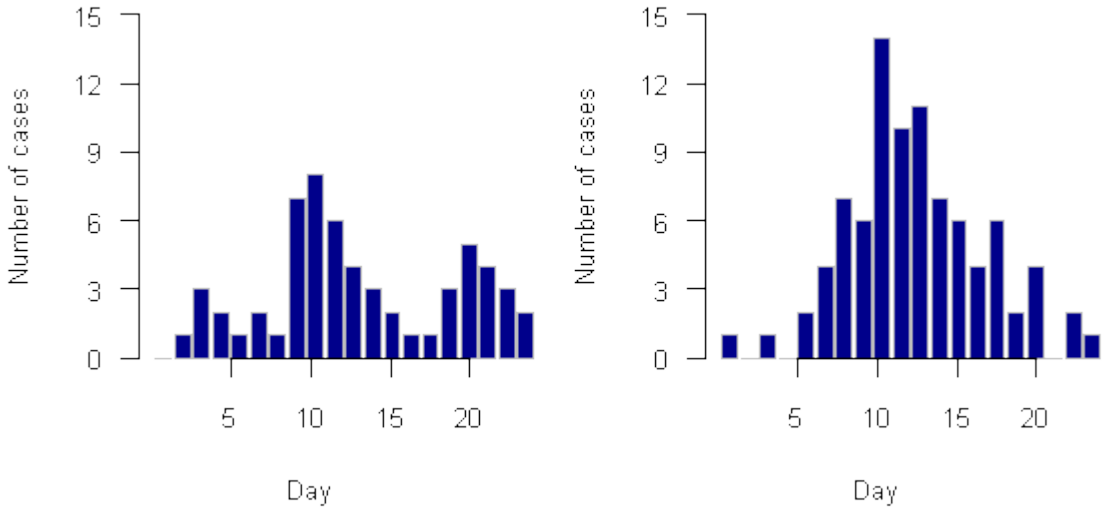


Figure 3: Epidemic curves. The plot on the left is typical of a propagated epidemic. The curve on the right is typical of a common source epidemic.

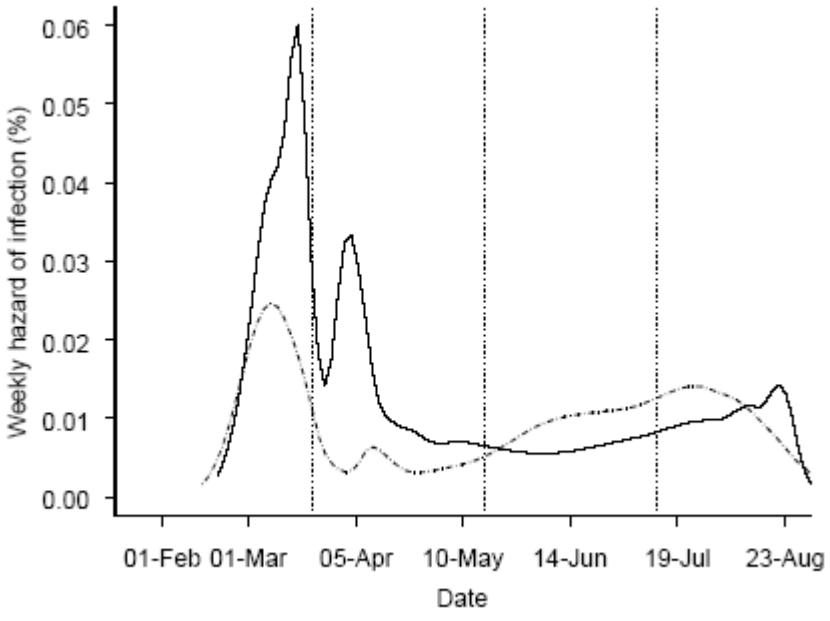


Figure 4: Weekly hazard of foot-and-mouth disease infection for cattle holdings (solid line) and 'other' holdings (dashed line) in Cumbria (Great Britain) in 2001. Reproduced from Wilesmith et al. (2003).

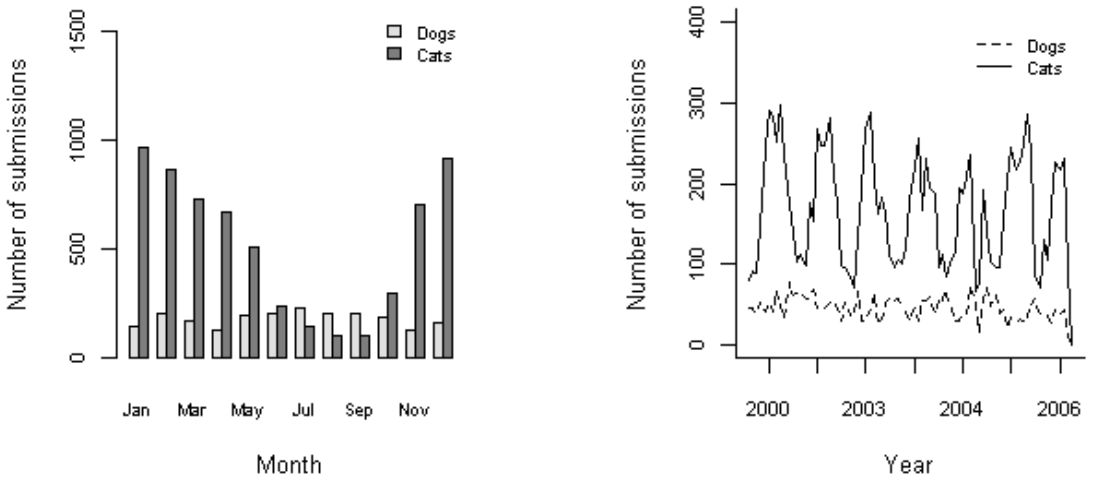


Figure 5: Free-roaming and surrendered dogs and cats submitted to a humane shelter in Wellington, New Zealand, 1999 – 2006 (Rinzin et al. 2008). The plot on the left shows the total number of dogs and cats submitted to the shelter per calendar month throughout the study period. The plot on the right shows monthly counts of submissions.

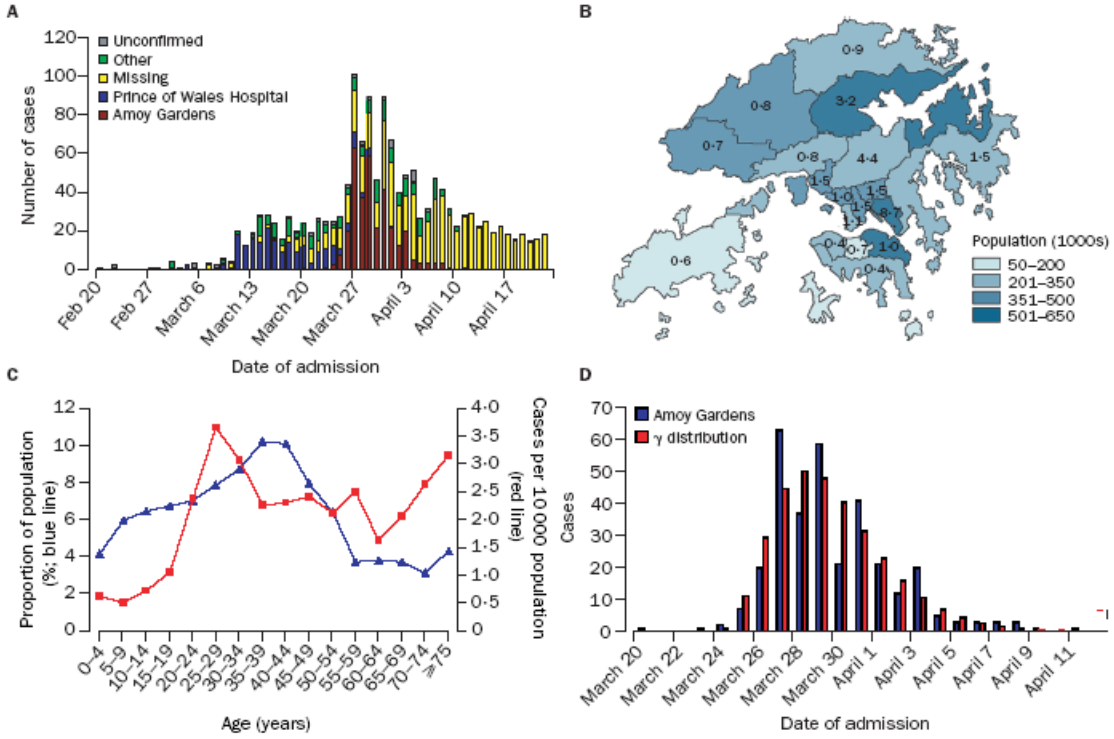


Figure 6: Descriptive epidemiology of Severe Acute Respiratory Syndrome in Hong Kong, February to April, 2003. A: Temporal pattern of SARS epidemic in Hong Kong by cluster of infection. B: Spatial distribution of population of Hong Kong and district-specific incidence (per 10 000 population) over course of epidemic to date. C: Age distribution of residents of Hong Kong and age-specific incidence (per 10,000 population) over course of epidemic to date. D: Detail of temporal pattern for Amoy Gardens cluster, according to day of admission, and fitted gamma distribution. Reproduced from Donnelly et al. (2004).

2 Measures of health

By the end of this unit you should be able to:

- Explain (with examples) why it is important to quantify the level of disease in a population.
- Explain the difference between prevalence and incidence, using examples.
- Describe the difference between incidence risk and incidence rate and explain when one measure might be preferred over the other.
- Describe the difference between closed and open populations, using examples.
- Calculate incidence risk and incidence rate for closed and open populations, given the appropriate data and formulae.
- Explain why adjusting disease frequency measures is useful in veterinary epidemiology, using examples.

A fundamental task in epidemiological research is to quantify the occurrence of disease. This can be done by counting the number of affected individuals. To compare levels of disease among groups of individuals, time frames and locations, we need to consider counts of cases in context of the size of the population from which those cases arose. Quantifying the levels of disease in a population is important since it allow animal health authorities to:

- Determine which diseases are of economic importance.
- Set priorities for the use of resources for disease control activities.
- Plan, implement and evaluate disease control programmes.
- Meet reporting requirements of international organisations (such as the Office International des Epizooties).
- Demonstrate disease freedom to trading partners.

Before discussing the methods for quantifying disease frequency it helps if we define some key terms.

A **proportion** is a fraction in which the numerator is included in the denominator. Say we have a herd of 100 cattle and over a 12-month period we identify 58 diseased animals. The proportion of diseased animals is $58 \div 100 = 0.58 = 58\%$.

A **ratio** defines the relative size of two quantities expressed by dividing one (numerator) by the other (denominator). The odds of disease (a ratio) in our herd of 100 cattle is 58:42 or 1.4 to 1.

A **closed population** is a populaton where no additions or removals occur during a defined follow-up period. An **open population** is a population where individuals are added (e.g. as births or purchases) and removed (e.g. as sales or deaths) during the follow-up period. Most of the populations that you will deal with in veterinary practice will be open.

The term **morbidity** is used to refer to the extent of disease or disease frequency within a defined population. Morbidity can be expressed as either **prevalence** or **incidence**.

2.1 Prevalence

Strictly speaking, the term prevalence refers to the number of cases of a given disease or attribute that exists in a population at a specified point in time. Prevalence risk is the proportion of a population that has a specific disease or attribute at a specified point in time. Many authors use the term 'prevalence' when they really mean prevalence risk, and these notes will follow this convention.

$$\text{Prevalence} = \frac{\text{Number of existing cases}}{\text{Size of population}} \tag{1}$$

Two types of prevalence are reported in the epidemiological literature: (1) **point prevalence** equals the number of disease cases in a population at a single point in time (a snapshot), (2) **period prevalence** equals the proportion of the population with a given disease or condition over a specific period of time. When calculating period prevalence the number of cases equals the number of individuals which have the disease at the start of the period plus the number of new cases that occur during the remainder of the follow-up period.

In 1944 the cities of Newburgh and Kingston, New York agreed to participate in a study of the effects of water fluoridation for prevention of tooth decay in children (Ast and Schlesinger 1956). In 1944 the water in both cities had low fluoride concentrations. In 1945, Newburgh began adding fluoride to its water — increasing the concentration ten-fold while Kingston left its supply unchanged. To assess the effect of water fluoridation on dental health, a survey was conducted among school children in both cities during the 1954 – 1955 school year. One measure of dental decay in children 6 – 9 years of age was whether at least one of a child's 12 deciduous cuspids or first or second deciduous molars was missing or had clinical or X-ray evidence of tooth decay.

Of the 216 first-grade children examined in Kingston, 192 had evidence of tooth decay. Of the 184 first-grade children examined in Newburgh 116 had evidence of tooth decay. Assuming complete survey coverage, there were 192 prevalent cases of tooth decay among first-grade children in Kingston at the time of the study. The prevalence of tooth decay was $192 \div 216 = 89$ cases per 100 children in Kingston and $116 \div 184 = 63$ cases per 100 children in Newburgh.

Reference: Ast DB, Schlesinger ER (1956). The conclusion of a ten-year study of water fluoridation. American Journal of Public Health, 46: 265-271.

2.2 Incidence

Incidence measures how frequently initially susceptible individuals become disease cases as they are observed over time. An incident case occurs when an individual changes from being susceptible to being diseased. The count of incident cases is the number of such events that occur in a defined population during a specified time period. There are two ways to express incidence: **incidence risk** and **incidence rate**.

Incidence risk

Incidence risk (also known as cumulative incidence) is the proportion of initially susceptible individuals in a population who become new cases during a defined follow-up period.

$$\text{Incidence risk} = \frac{\text{Number of incident cases}}{\text{Number of individuals initially at risk}} \tag{2}$$

Incidence risk is reported as the number of cases of disease per head of population over a specified follow-up period. The follow-up period may be arbitrarily fixed (e.g. the 5-year incidence risk of arthritis) or it may vary among individuals (e.g. the lifetime incidence risk of arthritis). In an investigation of a localised epidemic the follow-up period may be simply defined as the duration of the epidemic.

Last year a herd of 121 cattle were tested for tuberculosis using the tuberculin test and all tested negative. This year the same 121 cattle were tested and 25 tested positive.

The incidence risk of tuberculosis in this herd was 21 cases per 100 cattle for the 12-month follow-up period.

Calculating incidence risk for closed populations is straightforward. The denominator is simply the number of disease free individuals present at the start of the follow-up period.

For open populations things are a little more complicated: we need to take into account those individuals that enter and leave the population throughout the follow-up period. To do this we take the number of individuals present at the start, add half of the number that enter the population during the follow-up period (e.g. births and purchases) and subtract half the number that are lost (i.e. individuals that leave the population for reasons unrelated to the disease of interest). In effect this gives the population size at the mid-point of the follow-up period (assuming individuals enter and exit the population at a constant rate). If an animal can only experience one episode of disease we include diseased animals with the group that leave (i.e. once they've become a case they are removed from the population at risk).

- Number at risk = population size at the mid-point of the follow-up period.
- Number at risk = $[N_{start} + \frac{1}{2}N_{new}] - [\frac{1}{2}N_{lost}]$.
- Number at risk = $[N_{start} + \frac{1}{2}N_{new}] - [\frac{1}{2}(N_{lost} + N_{cases})]$. This approach assumes that only one case of disease is considered per individual.

Correctly estimating the size of the population at risk (the denominator) presents the most difficulties when calculating incidence. Remember the following rules of thumb:

- If the population is closed the population at risk equals the number of disease free individuals present at the start of the follow-up period.
- If the population is open the population at risk should be adjusted to account for those that enter and leave the population throughout the follow-up period.

You conduct a prospective observational study to document the incidence of a disease in a population of cattle. At the start of 1000 healthy animals are enrolled. During the first month of the study 44 head of cattle are purchased and are added to the study population, 20 animals are removed (for reasons unrelated to the disease of interest) and 112 cattle are diagnosed with the disease of interest. Assuming that each animal can only become diseased once, what is the incidence risk of disease in this population during the first month?

N_{start} = 1000, N_{new} = 44, N_{lost} = 20, N_{cases} = 112

Number at risk = $[1000 + (\frac{1}{2} \times 44)] - [\frac{1}{2}(20 + 112)]$
Number at risk = 1022 - 66
Number at risk = 956

The incidence risk of disease was 112 cases per 956 cattle (equivalent to 12 cases per 100 cattle) during the 1-month follow-up period.

Incidence rate

Incidence rate (also known as incidence density) is the number of new cases of disease that occur per unit of individual time at risk during a defined follow-up period.

$$\text{Incidence rate} = \frac{\text{Number of incident cases}}{\text{Amount of at-risk experience}} \tag{3}$$

Because the denominator is expressed in units of animal- or person-time at risk those individuals that are withdrawn or are lost to follow-up are easily accounted-for. Consider a study of clinical mastitis in five cows over a 12-month period, as shown in Table 1.

Table 1: Hypothetical mastitis data

ID	Details	Events	Days at risk
1	Calve 01 Aug, mastitis 15 Aug, mastitis 15 Sep, mastitis 15 Oct, sold 15 Nov	3	106 [a]
2	Calve 01 Aug, mastitis 15 Nov, dry off 15 May,	1	365
3	Purchased 01 Dec, mastitis 01 Jan, Dry off 15 May	1	243
4	Calve 01 Aug, Sold 16 Nov	0	107
5	Calve 01 Oct, Died 05 Oct	0	4
Total		5	825

[a] 15 Nov 2001 - 01 Aug 2001 = 106 days.

On the basis of the data presented in Table 1 the incidence rate of clinical mastitis for the 12-month period is 5 cases per 825 cow-days at risk (equivalent to 2.2 cases of clinical mastitis per cow-year at risk).

For closed populations the amount of at-risk experience (the denominator) is the number of disease free individuals present at the start of the follow-up period multiplied by the length of the follow-up period. For open populations we take the number of individuals present at the start, add half of the number that enter the population during the follow-up period and subtract half the number that leave (just as we did when calculating incidence risk for an open population). This number — effectively the population size at the mid-point of the follow-up period — is then multiplied by the length of the follow-up period to provide an estimate of the total at-risk experience.

- At-risk experience = population size at the mid-point of the study period $\times$ length of study period.
- At-risk experience = $\{[N_{start} + \frac{1}{2}N_{new}] - [\frac{1}{2}N_{lost}]\} \times$ length of study period.
- At-risk experience = $\{[N_{start} + \frac{1}{2}N_{new}] - [\frac{1}{2}(N_{lost} + N_{cases})]\} \times$ length of study period. This approach assumes that only one case of disease is considered per individual.

Herd management software packages should be able to calculate the exact amount of at-risk experience because the date of entry and exit is known for each individual member of the population and it is a simple job to sum the at-risk experience for each individual to yield the total at-risk experience for the population. The method described here should be used when you want to estimate incidence rate on the basis of summary data (i.e. when the only information you have is the total number of animals present at the start of the follow-up period, the total number of additions and the total number of removals).

Gardner et al. (1999) studied on-the-job back sprains and strains among 31,076 material handlers employed by a large retail merchandising chain. Payroll data for a 21-month period during 1994 – 1995 were linked with job injury claims. A total of 767 qualifying back injuries occurred during 54,845,247 working hours, yielding an incidence rate of 1.40 back injuries per 100,000 worker-hours.

Reference: Gardner LI, Landsittel DP, Nelson NA (1999). Risk factors for back injury in 31,076 retail merchandise store workers. American Journal of Epidemiology, 150: 825 - 833.

The relationship between prevalence and incidence

Table 2 compares the main features of the three measures of disease frequency we have defined. Figure 7 provides a worked example. This example is based on a herd of 10 animals which are all disease-free at the beginning of the observation period and followed for a 12-month period. Disease status is assessed at monthly intervals.

The relationship between point prevalence, period prevalence and incidence can be explicated using an analogy with photography. Point prevalence is like a flashlit photograph: what is happening at an instant in time. Period prevalence is analogous to a long exposure: the number of events recorded in the photo whilst the camera shutter was open. In a movie each frame records an instant (point prevalence). By looking from frame to frame one notices new events (incident events) and can relate the number of such events to a period (number of frames) to produce incidence rate.

Non-epidemiologists tend to have difficulty in understanding incidence rate so in some situations it may be useful to convert incidence rate estimates to incidence risk (so you're talking about the number of cases of disease per head of population). Providing incidence rate is constant throughout the follow-up period, incidence risk equals:

- Closed population: incidence risk = incidence rate × length of study period.
- Open population (when the study period is short): incidence risk $\sim$ incidence rate × length of study period.
- Open population: incidence risk $= 1 - \exp^{(\text{-incidence rate} \times \text{length of study period})}$

Prevalence can be estimated from incidence rate, providing incidence rate is constant throughout the follow-up period and the population is closed:

- Prevalence = (incidence rate × duration of disease) ÷ (incidence rate × duration of disease + 1).
- Duration of disease = (prevalence) ÷ (incidence rate × 1 - prevalence).

Table 2: A comparison of the main features of prevalence, incidence risk, and incidence rate.

	Point prevalence	Period prevalence	Incidence risk	Incidence rate
Numerator	All cases counted on a single occasion	Cases present at period start + new cases during follow-up period	New cases during follow-up period	New cases during follow-up period
Denominator	All individuals examined	All individuals examined	All susceptible individuals present at the start of the study	Sum of time period at risk for susceptible individuals present at the start of the study
Time	Single point or period	Defined follow-up period	Defined follow-up period	Measured for each individual from beginning of study until disease event, exit from the population, or end of the follow-up period
Study type	Cross-sectional	Cohort	Cohort	Cohort
Interpretation	Probability of having disease at a given point in time	Probability of having disease over a defined follow-up period	Probability of developing disease over a defined follow-up period	How quickly new cases develop over a defined follow-up period

In a herd of dairy cows the incidence rate of lameness is estimated to be 0.006 cases per cow-day at risk. The average duration of disease is 7 days.

The estimated prevalence of disease is $(0.006 \times 7) \div (0.006 \times 7 + 1) = 0.041$, that is 4.1 cases per 100 cows.

2.3 Other measures of health

Attack rates

Attack rates are usually used in outbreak situations where the period of risk is limited and all cases arising from exposure are likely to occur within the risk period. Attack rate is defined as the number of cases divided by the number of individuals exposed. 'Attack risk' would be a better term for this parameter.

Secondary attack rates

Secondary attack rates are used to describe infectiousness. The assumption is that there is spread of an agent within an aggregation of individuals (e.g. a herd or family) and that not all cases are a result of a common-source exposure. Secondary attack rates are the number of cases at the end of the study period less the number of initial (primary) cases divided by the size of the population that were initially at risk. Again, 'secondary attack risk' would be a better term for this parameter.

Mortality

Mortality risk (or rate) is an example of incidence where death is the outcome of interest. Cause-specific mortality risk is the incidence risk of fatal cases of a particular disease in a population

Animal	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Diseased?	Months at risk
A					Disease								yes	4
B													no	12
C								Withdrawn					no	7
D		Disease											yes	1
E													no	12
F						Disease							yes	5
G											Disease		yes	10
H													no	12
I													no	12
J						Withdrawn							no	5
Total													4	80

Number of disease events:	4	Number present at start:	10
		Number of withdrawals:	2
		Number present at end:	8
Prevalence in June:		33% (3 cases in 9 animals)	
Prevalence in December:		50% (4 cases in 8 animals)	
Incidence risk (accounting for withdrawals):		44% (4 cases in 9 animals)	
Incidence risk (approximate):		40% (4 cases in 10 animals)	
Incidence rate (exact):		4 cases per 80 animal-months at risk	
Incidence rate (approximate):		4 cases per 84 animal-months at risk	

Figure 7: Calculation of prevalence, incidence risk and incidence rate (using exact and approximate methods).

at risk of death from that disease. The denominator includes both prevalent cases of the disease (that is, the individuals that haven't died yet) as well as individuals who are at risk of developing the disease.

Case fatality

Case fatality risk (or rate) refers to the incidence of death among individuals who develop the disease.

Case fatality risk reflects the prognosis of disease among cases, while mortality reflects the burden of deaths from the disease in the population as a whole.

Proportional mortality

As its name implies, proportional mortality is the proportion of all deaths that are due to a particular cause for a specified population and time period:

$$\text{Proportional mortality} = \frac{\text{Number of deaths from the disease}}{\text{Number of deaths from all causes}} \tag{4}$$

2.4 Adjusted measures of health

Often, we want to compare the frequency of disease in different populations (e.g. herds, regions, countries). However, since disease frequency often depends on age, a higher incidence of disease in one population may simply be due to the fact that it is generally older than a second population. To avoid this problem we can standardise disease frequency estimates, effectively eliminating the effect of age. Disease frequency estimates computed using these techniques are referred to as age-adjusted or age-standardised (note that we can adjust on the basis of other variables — for example by sex, herd type, and region).

There are two methods for adjusting disease frequency estimates: **direct adjustment** and **indirect adjustment**.

Direct adjustment

With direct adjustment the adjusted count for the i^{th} strata equals the observed disease frequency estimate (i.e. prevalence or incidence) multiplied by a standard population estimate for the i^{th} strata:

$$\text{Directly adjusted count}_i = \text{OBS R}_i \times \text{STD P}_i \tag{5}$$

Where:

OBS R_i: the observed prevalence or incidence in the i^{th} strata
STD P_i: the size of the standard population in the i^{th} strata

If we were adjusting on the basis of sex we might say that in a standard population 50% of the total population would be allocated to the male strata and 50% to the female strata. The choice of the standard population for direct adjustment is not crucial, however, where possible it is desirable to select a standard that is demographically sensible. Consider a study of leptospirosis seroprevalence in Scottish dogs, the details of which are shown in Table 3.

Table 3: Seroprevalence of leptospirosis in urban dogs, stratified by city.

City	Positive	Sampled	Seroprevalence
Edinburgh	61	260	23%
Glasgow	69	251	27%
Total	130	511	25%

The crude data suggests that Glasgow has a slightly higher seroprevalence of leptospirosis amongst its dog population. However, what about the sex composition of the two populations that were studied? Male dogs are known to have a higher incidence rate for leptospirosis

because of their sexual behaviour, and it might be that more male dogs were sampled in Glasgow. Sex-specific prevalence estimates confirm the role of population structure (Table 4).

Table 4: Seroprevalence of leptospirosis in urban dogs, stratified by city and sex.

City	Positive		Sampled		Seroprevalence		
	Male	Female	Male	Female	Male	Female	Total
Edinburgh	15	46	48	212	31%	22%	23%
Glasgow	53	16	180	71	29%	22%	27%
Total	68	62	228	223	30%	22%	25%

The confounding effect of sex can be removed by producing gender-adjusted prevalence estimates (Table 5). Direct adjustment involves, for each strata, mutliplying the crude seroprevalence estimates by a standard population estimate. The sum of the directly adjusted counts across all strata divided by the size of the standard population provides the directly adjusted disease frequency estimate for the population. In this example, we use a standard population comprised of 250 males and 250 females.

Table 5: Directly adjusted seroprevalence of leptospirosis in urban dogs, stratified by city.

City	Positive		Sampled		Seroprevalence
	Male	Female	Male	Female	
Edinburgh	0.31×250=77	0.22×250=55	250	250	(77 + 55) / 500 = 26%
Glasgow	0.29×250=72	0.22×250=55	250	250	(72 + 55) / 500 = 25%
Total	77+72=149	55+55=110	500	250	(149 + 110) / 1000 = 25%

The directly adjusted prevalence estimates are similar which suggests the difference between the cities is due to the different sex structures of the two populations.

Indirect adjustment

With indirect adjustment the adjusted count for the i^{th} strata equals a standardised frequency estimate multiplied by the observed population size for the i^{th} strata:

$$\text{Indirectly adjusted count}_i = \text{STD R}_i \times \text{OBS P}_i \tag{6}$$

Where:

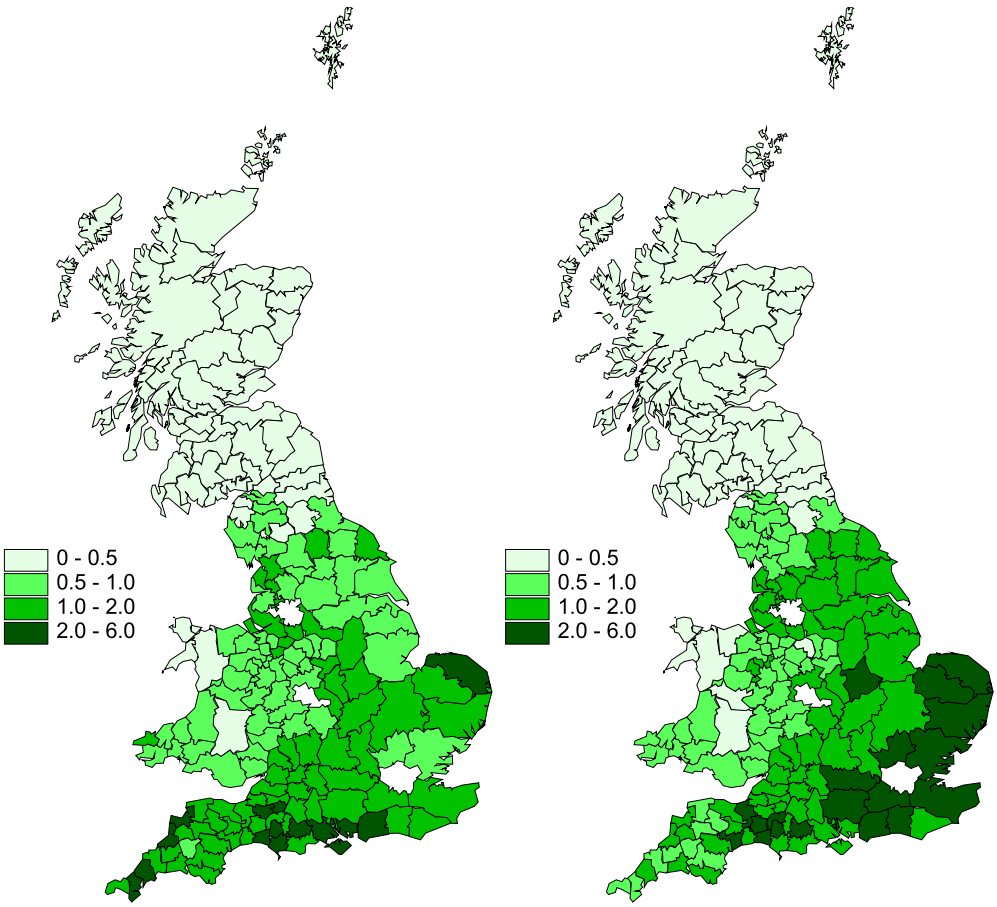
STD R_i: the standard incidence or prevalence in the i^{th} strata of the population
OBS P_i: the observed population size in the i^{th} strata

It is usual to set the standardardised incidence (or prevalence) for the i^{th} strata as the sum of the total number of disease events across all strata divided by the total population size. Using

this approach the indirectly adjusted disease count for the i^{th} strata (also known as the expected number of disease events, E_i) equals the standardardised incidence (or prevalence) multiplied by the population size for each strata.

It is common to divide the observed number of disease events (O_i) per strata by the expected number (E_i) to yield a standardised morbidity or mortality ratio (SMR). If area units (e.g. states, counties, census tracts) are the basis for stratification it is common to plot the SMR for each area unit i in the form of a choropleth map (a map where areas are coloured according to the value of the outcome of interest). Choropleth maps of SMR estimates are an effective way to describe the geographical distribution of disease in a population, and how this might change over time (Figure 8).

We know that the prevalence of a given disease throughout a country is 0.01%. If we are presented with a region with 20,000 animals the expected number of cases of disease in this region will be $0.0001 \times 20{,}000 = 2$. If the actual number of cases of disease in this region is 5, then the standardised mortality (morbidity) ratio is $5 \div 2 = 2.5$. That is, there were 2.5 times more cases of disease in this region, compared with the number of cases expected.



(a) SMR: pre-control cohort

(b) SMR: post-control cohort

Figure 8: An example of the use of indirect standardisation used to describe the change in spatial distribution of disease risk over time. Choropleth maps of area-level standardised mortality ratios (SMRs) for bovine spongiform encephalopathy in British cattle 1986 – 1997, for (a) cattle born before the 18 July 1988 ban on feeding meat and bone meal to ruminants, and (b) cattle born between 18 July 1988 and 30 June 1997. The above maps show a shift in area-level risk over time towards the east of the country (even though the incidence of BSE reduced markedly from 1988 to 1997). Reproduced from Stevenson et al. (2005).

3 Measures of association

By the end of this unit you should be able to:

- Given disease count data, construct a 2 × 2 table and, given the appropriate formulae, explain how to calculate the following measures of association: risk ratio, odds ratio, attributable risk, attributable fraction, population attributable risk, population attributable fraction.
- Interpret the following measures of association: risk ratio, odds ratio, attributable risk, attributable fraction, population attributable risk, population attributable fraction.
- Describe situations where the risk ratio is not a valid measure of association between exposure and outcome.

Risk is the probability that an event will happen. A characteristic or factor that influences whether or not an event occurs, is called a risk factor.

- Worn tyres are a risk factor for motor vehicle accidents.
- High blood pressure is a risk factor for coronary heart disease.
- Vaccination is a protective risk factor in that it usually reduces the risk of disease.

If we identify those risk factors that are causally associated with an increased likelihood of disease and those causally associated with a decreased likelihood of disease, then we are in a good position to make recommendations about health management. Much of epidemiological research is concerned with identifying and quantifying the effect of risk factors on the likelihood of disease.

Associations between putative risk factors (exposures) and an outcome (usually a disease) can be investigated using analytical observational studies. Consider a study where subjects are disease free at the start of the study and all are monitored for disease occurrence for a specified time period. If both exposure and outcome are binary variables (yes or no), the results can be presented as a 2 × 2 table.

	Diseased	Non-diseased	Total
Exposed	a	b	$a+b$
Non-exposed	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d=n$

Based on data presented in this 'standard' format (i.e. disease status shown in the columns and exposure status shown in the rows), various measures of association can be calculated. These fall into three main categories: (1) measures of strength, (2) measures of effect, and (3) measures of total effect. To calculate these parameters, it helps to first work out some summary parameters:

Incidence risk in the exposed population: $R_E = a/(a+b)$
Incidence risk in the non-exposed population: $R_O = c/(c+d)$
Incidence risk in the total population: $R_{total} = (a+c)/(a+b+c+d)$
Odds of disease in the exposed population: $O_E = a/b$
Odds of disease in the non-exposed population: $O_O = c/d$

Observed associations are not always causal and/or may be estimated with bias. The interpretation of the following measures of association assumes that relationships are causal and have been estimated without bias.

3.1 Measures of strength

Incidence risk ratio

The incidence risk ratio is defined as the incidence risk of disease in the exposed group divided by the incidence risk of disease in the unexposed group:

$$RR = \frac{R_E}{R_O} \tag{7}$$

The incidence risk ratio provides an estimate of how many times more likely exposed individuals are to experience disease, compared with non-exposed individuals. If the incidence risk ratio equals 1, then the risk of disease in the exposed and non-exposed groups are equal. If the incidence risk ratio is greater than 1, then exposure increases the risk of disease with greater departures from 1 indicative of a stronger effect. If the incidence risk ratio is less than 1, exposure reduces the risk of disease and exposure is said to be protective. The incidence risk ratio cannot be estimated in case-control studies, as these studies do not allow calculation of risks. Odds ratios are used instead — see below.

Risk ratios range between 0 and infinity.

Incidence rate ratio

In a study where incidence rate has been measured (rather than incidence risk) the incidence rate ratio (also known as the rate ratio) can be calculated. This is the ratio of the incidence rate in the exposed group to that in the non-exposed group. The incidence rate ratio is interpreted in the same way as the incidence risk ratio.

The term **relative risk** is used as a synonym for both incidence risk ratio and incidence rate ratio.

Odds ratio

Where odds ratio is defined as the odds of disease in the exposed group divided by the odds of disease in the unexposed group. The odds ratio (OR) is an estimate of incidence risk ratio and is interpreted in the same way. The odds ratio is calculated as:

$$OR = \frac{O_E}{O_O} = \frac{ad}{bc} \tag{8}$$

When the number of cases of disease is low relative to the number of non-cases (i.e. the disease is rare), then the odds ratio will approximate the incidence risk ratio. If the incidence of disease is relatively low in both exposed and non-exposed individuals, then a will be small relative to b and c will be small relative to d. As a result:

$$RR = \frac{a/(a+b)}{c/(c+d)} \simeq \frac{a/b}{c/d} = \frac{ad}{bc} = OR \tag{9}$$

3.2 Measures of effect in the exposed population

Attributable risk (rate)

Attributable risk (or rate) is defined as the increase or decrease in the risk (or rate) of disease in the exposed group that is attributable to exposure. Attributable risk (unlike incidence risk ratio) measures the absolute quantity of the outcome measure that is associated with the exposure. Using the notation defined above, attributable risk (AR) is calculated as:

$$AR = R_E - R_O \tag{10}$$

In a clinical setting attributable risk may also be referred to as attributable risk reduction (ARR) or attributable risk increase (ARI) depending on whether the event risk is decreased or increased in the exposure positive (treatment) group.

Another useful way of expressing attributable risk in a clinical setting is in terms of the number needed to treat (NNT). The NNT is the number of subjects who would have to be given the exposure (i.e. treatment) to prevent a negative outcome from occurring. NNT equals the inverse of the attributable risk.

A prospective cohort study was conducted to evaluate the effect of administering oxygen to patients with renal impairment prior to general anaesthesia. The incidence risk of death in oxygen treated patients was 3.5 cases per 100. The incidence risk of death in patients not receiving oxygen was 6.7 cases per 100. The attributable risk was 3.5 - 6.7 = -3.2 cases per 100. In other words, oxygen treatment prevented death in 3.2% of patients. The NNT for these data was -31.3. This means that around 31 patients would need to be treated with oxygen to prevent one death.

NNT gives a good intuitive feel for the treatment benefit and is often useful when communicating the results of such studies to clients.

Attributable fraction

Attributable fraction (also known as the attributable proportion in exposed subjects) is the proportion of disease in the exposed group that is due to exposure. Using the notation defined above, attributable fraction (AF) is calculated as:

$$AF = \frac{(R_E - R_O)}{R_E} = \frac{(RR - 1)}{RR} \tag{11}$$

For case-control studies, attributable fraction can be estimated if the incidence of disease is low:

$$AF_{est} = \frac{(O_E - O_O)}{O_E} = \frac{(OR - 1)}{OR} \tag{12}$$

In vaccine trials, vaccine efficacy is defined as the proportion of disease prevented by the vaccine in vaccinated individuals (equivalent to the proportion of disease in unvaccinated individuals due to not being vaccinated), which is the attributable fraction.

A case-control study investigating the effect of oral vaccination on the presence or absence of rabies in foxes was conducted. The results shown in Table 6 were obtained.

The odds of rabies in the unvaccinated group was 2.3 times the odds of rabies in the vaccinated group (OR = 2.30). Fifty six percent of rabies cases in unvaccinated foxes was due to not being vaccinated ($\text{AF}_{\text{est}} = 0.56$).

Table 6: Oral vaccination and the risk of rabies in wild foxes.

	Rabies +	Rabies -	Total
Vaccination -	18	30	48
Vaccination +	12	46	58
Total	30	76	106

3.3 Measures of effect in the total population

Population attributable risk (rate)

Population attributable risk (or rate) is the increase or decrease in risk (or rate) of disease in the population that is attributable to exposure. Using the notation defined above, population attributable risk (PAR) is calculated as:

$$PAR = R_{total} - R_O \tag{13}$$

Population attributable fraction

Population attributable fraction (also known as the aetiologic fraction) is the proportion of disease in the population that is due to the exposure. Using the notation defined above, the population attributable fraction (PAF) is calculated as:

$$PAF = \frac{(R_{total} - R_O)}{R_{total}} \tag{14}$$

Methods are available to estimate PAF using data from case-control studies.

A cohort study investigated the relationship between dry cat food (DCF) and feline urologic syndrome (FUS). The results shown in Table 7 were obtained.

Table 7: Use of dry cat food and the presence of FUS in cats.

	FUS +	FUS -	Total
DCF +	13	2163	2176
DCF -	5	3349	3354
Total	18	5512	5530

The incidence risk of FUS in the DCF+ group was 5.97 cases per 1000. The incidence risk of FUS in the DCF group was 1.49 cases per 1000. The incidence risk of FUS in DCF exposed cats was 4.01 times greater than the incidence risk of FUS in DCF cats.

The incidence risk of FUS in DCF+ cats that may be attributed to DCF is 4.5 per 1000 (AR = 0.0045). In DCF+ cats 75% of FUS is attributable to DCF (AF = 0.75).

The incidence risk of FUS in the cat population that may be attributed to DCF is 1.8 per 1000. That is, we would expect the risk of FUS to decrease by 1.8 cases per 1000 if DCF were not fed (PAR = 0.0018). Fifty-four percent of FUS cases in the cat population are attributable to DCF (PAF = 0.54).

3.4 Using the appropriate measure of effect

Table 8 outlines which measures of effect are appropriate for each of the three major study designs (case-control, cohort and cross-sectional studies).

Table 8: Epidemiologic measures of association for independent proportions in 2 × 2 tables.

Parameter	Case-control	Cohort	Cross-sectional
Measures of strength:			
Incidence risk ratio	No	Yes	Yes (prevalence RR)
Incidence rate ratio	No	Yes	No
Odds ratio	Yes	Yes	Yes (prevalence OR)
Measures of effect:			
Attributable risk	No	Yes	Yes
Attributable fraction	No	Yes	Yes
Attributable fraction (est)	Yes	Yes	Yes
Measures of effect in population (total effect):			
Population attributable risk	No	Yes[a]	Yes
Population attributable fraction	No	Yes[a]	Yes
Population attributable fraction (est)	Yes	Yes	Yes

[a] If an estimate of the prevalence of exposure or disease incidence for the population is available from another source.

Members of the public often have a poor understanding of relative and absolute risk. A case in point was a recent news item describing the results of a study of risk factors for leukaemia in children (Draper et al. 2005). Children who lived within 200 metres of high voltage lines at birth had a 70% higher incidence risk of leukaemia compared with those that lived 600 metres or more away. While the facts were correctly reported, the interpretation of the scientific evidence was misguided. If the incidence risk of leukaemia in the general population is around 1 in 20,000 a 70% increase elevates this to around 2 cases per 20,000 (a very minor increase in absolute terms).

Pylon study sparks child health fears

Large-scale study finds increased incidence of leukaemia in children who live near high-voltage lines.

KELLY ANDREW and NIKKI MACDONALD

FEARS that overhead power lines could increase the risk of childhood leukaemia have been heightened by a British study.

In one of the biggest studies of its kind, Oxford University researchers investigated more than 29,000 children with cancer who were born between 1962 and 1995, including 9700 with leukaemia (a cancer of the blood), compared with a control group of healthy children.

Youngsters who lived within 200 metres of high-voltage lines at birth had a 70 per cent higher incidence of leukaemia than those 600 metres or more away. Those living between 200 metres and 600 metres from the lines were 23 per cent more likely to develop leukaemia.

Unfazed: Rosemary Greenfield says her children Jamie, 11, and Belinda, 9, have lived near a pylon all their lives and are healthy.

lions and child cancer, rather than resolving it.

During the past 20 years, scien-

and Belinda, 9, for 12 years.

She was not rattled by the findings of the British study. "I was a

Figure 9: Newspaper headline warning of the risk of leukaemia associated with living close to high-voltage electricity lines. Source: The Dominion Post (Wellington, New Zealand) Saturday 4 June 2005.

4 Study design

By the end of this unit you should be able to:

- Describe the difference between descriptive and analytical epidemiological studies (giving examples of each).
- Describe the major features of the following study designs: case reports, case series, descriptive studies, ecological studies, cross-sectional studies, cohort studies, case-control studies, clinical trials, randomised clinical trials, and community trials.
- Suggest an appropriate study design to identify risk factors for disease, given details of a disease problem in a population of animals. Be able to justify your chosen design.
- Describe the strengths and weaknesses of cross-sectional studies, cohort studies, case-control studies, and clinical trials.

A study generally begins with a research question. Once the research question has been specified the next step is to choose a study design. A study design is a plan for selecting study subjects and for obtaining data about them. Figure 10 shows the major types of epidemiological study designs. There are three main study types: (1) descriptive studies, (2) analytical studies, and (3) experimental studies.

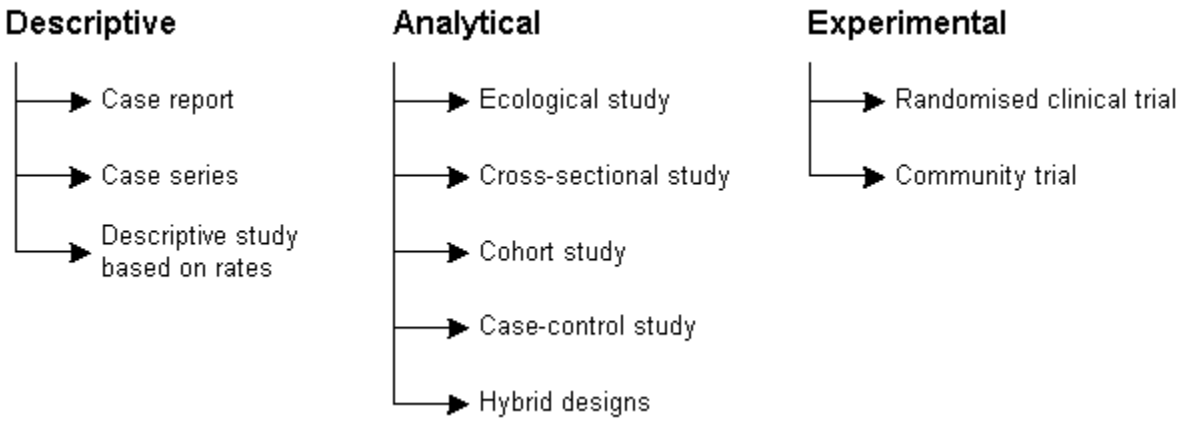


Figure 10: Tree diagram outlining relationships between the major types of epidemiologic study designs.

Descriptive studies are those undertaken without a specific hypothesis. They are often the earliest studies done on a new disease in order to characterise it, quantify its frequency, and determine how it varies in relation to individual, place and time. Analytical studies are undertaken to identify and test hypotheses about the association between an exposure of interest and a particular outcome. Experimental studies are also designed to test hypotheses between specific exposures and outcomes — the major difference is that in experimental studies the investigator has direct control over the study conditions.

4.1 Descriptive studies

The hallmark of a descriptive study is that it is undertaken without a specific hypothesis.

Case reports

A case report describes some 'newsworthy' clinical occurrence, such as an unusual combination of clinical signs, experience with a novel treatment, or a sequence of events that may suggest previously unsuspected causal relationships. Case reports are generally reported as a clinical narrative.

Trivier at al (2001) reported the occurrence of fatal aplastic anaemia in an 88 year-old man who had taken clopidogrel, a relatively new drug on the market that inhibits platelet aggregation. The authors speculated that his fatal illness may have been caused by clopidogrel and wished to alert other clinicians to a possible adverse effect of the drug.

Reference: Trivier JM, Caron J, Mahieu M, Cambier N, Rose C (2001). Fatal aplastic anaemia associated with clopidogrel. Lancet, 357: 446.

Cases series

Whereas a case report shows that something can happen once, a case series shows that it can happen repeatedly. A case series identifies common features among multiple cases and describes patterns of variability among them.

After bovine spongiform encephalopathy (BSE) appeared in British cattle in 1987, there was concern that the disease might spread to humans. A special surveillance unit was set up to study Creutzfeld-Jacob disease (CJD), a rare and fatal progressive dementia that shares clinical and pathological features of BSE. In 1996 investigators at the unit described ten cases that met the criteria for CJD but had all occurred at unusually young ages, showed distinctive symptoms and, on pathological examination, had extensive prion protein plaques throughout the brain similar to BSE.

Reference: Will RG, Ironside JW, Zeidler M, Cousens SN, Estibeiro K, Alperovitch A (1996). A new variant of Creutzfeld-Jacob disease in the UK. Lancet, 347: 921 - 925.

Descriptive studies based on rates

Descriptive studies based on rates quantify the burden of disease on a population using incidence, prevalence, mortality or other measures of disease frequency. Most use data from existing sources (such as birth and death certificates, disease registries or surveillance systems). Descriptive studies can be a rich source of hypotheses that lead later to analytic studies.

Schwarz et al. (1994) conducted a descriptive epidemiological study of injuries in a predominantly African-American part of Philadelphia. An injury surveillance system was set up in a hospital emergency centre. Denominator information came from US census data. These authors found a high incidence of intentional interpersonal injury in this area of the city.

Reference: Schwarz DF, Grisso JA, Miles CG, Holmes JH, Wishner AR, Sutton RL (1994). A longitudinal study of injury morbidity in an African-American population. Journal of the American Medical Association, 271: 755 - 760.

4.2 Analytical studies

Analytical studies are undertaken to test a hypothesis. In epidemiology the hypothesis typically concerns whether a certain **exposure** causes (or is assoicated with) a certain **outcome** — e.g. does cigarette smoking cause lung cancer? The term exposure is used to refer to any trait, behaviour, environmental factor or other characteristic as a possible cause of disease. Synonyms for exposure are: potential risk factor, putative cause, independent variable, and predictor. The term outcome generally refers to the occurrence of disease. Synonyms for outcome are: effect, end-point, and dependent variable.

The hypothesis in an analytic study is whether an exposure actually causes an outcome (not merely whether the two are associated). Each of Hill's criteria for causation are usually required to be met to support a case for causality, but probably the most important is that exposure must precede the outcome in time.

Ecological studies

In an ecological study the unit of analysis is a group of individuals (such as counties, states, cities, or census tracts). Summary measures of exposure and summary measures of outcome are compared and inference is made at the individual level. Ecological studies are relatively quick and inexpensive to perform and can provide clues to possible associations between exposures and outcomes of interest. A major disadvantage of ecological studies is that of ecological fallacy: the assumption that an observed relationship in aggregated data will hold at the individual level.

Yang et al. (1998) conducted an ecological study examining the association between chlorinated drinking water and cancer mortality among 28 municipalities in Taiwan. The investigators found a positive association between the use of chlorinated drinking water and mortality from rectal, lung, bladder, and kidney cancer.

Reference: Yang CY, Chiu HF, Cheng MF, Tsai SS (1998). Chlorination of drinking water and cancer in Taiwan. Environmental Research, 78: 1 - 6.

Cross-sectional studies

In a cross-sectional study a sample of individuals from a population is taken at a point in time. Individuals included in the sample are examined for the presence of disease and their status with regard to the presence or absence of specified risk factors. Cross sectional studies commonly involve surveys to collect data. Surveys range from simple one-page questionnaires addressing a single variable, to highly complex, multiple page designs. There is a whole sub-field of epidemiology associated with design, implementation and analysis of questionnaires and surveys.

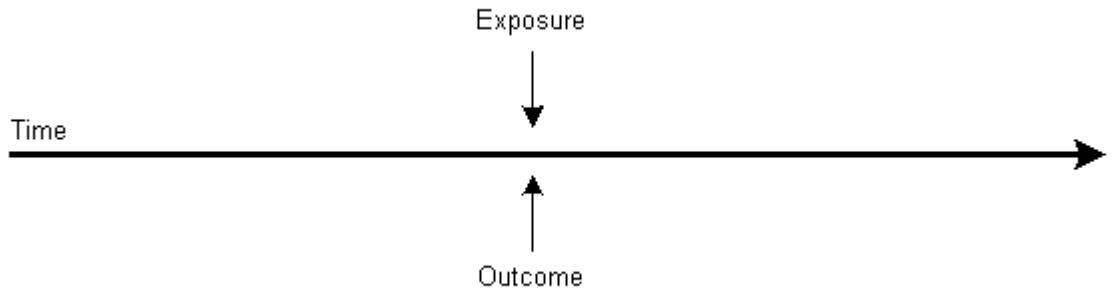


Figure 11: Schematic diagram of a cross-sectional study.

Advantages: Cross-sectional studies are relatively quick to conduct and their cost is moderate, compared with other study designs.

Disadvantages: Cross-sectional studies cannot provide information on the incidence of disease in a population — only an estimate of prevalence. Difficult to investigate cause and effect relationships.

Anderson et al. (1998) studied 4,063 children aged 8 to 16 years who had participated in the National Health and Nutrition Examination Survey to assess the relationship between television watching and body-mass index. At a single examination, each child was asked a series of questions about their usual amount of television viewing. Height, weight and a series of other body measurements were taken at the same time.

Boys and girls who reported watching four or more hours of television per day had significantly greater body mass indexes than boys and girls who reported watching fewer than two hours of television per day.

Reference: Anderson RE, Crespo CJ, Bartlett SJ, Cheskin LJ, Pratt M (1998). Relationship of physical activity and television watching with body weight and level of fatness among children. Results from the Third National Health and Nutrition Examination Survey. Journal of the American Medical Association, 279: 938 - 942.

Cohort studies

A cohort study involves comparing disease incidence over time between groups (cohorts) that are found to differ on their exposure to a factor of interest. Cohort studies are either **prospective** or **retrospective** (Figure 12).

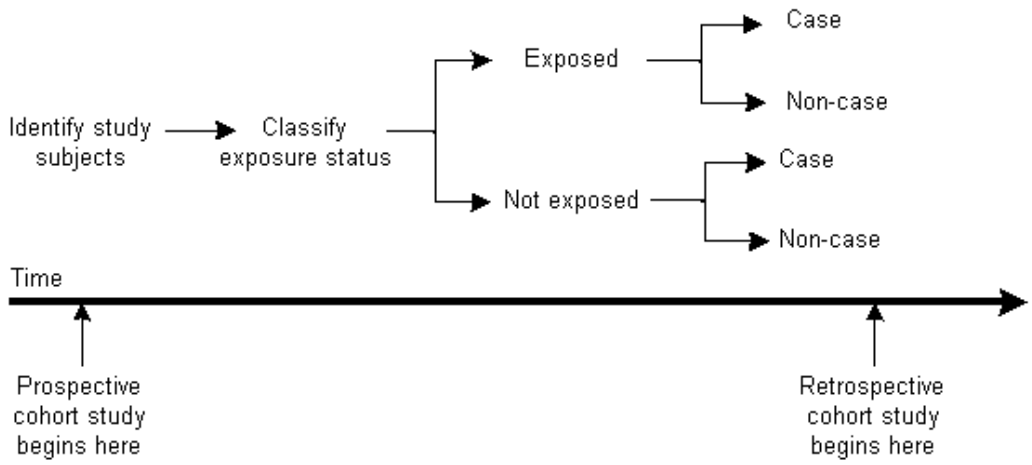


Figure 12: Schematic diagram of a prospective and retrospective cohort study.

A prospective cohort study begins with the selection of two groups of non-diseased animals, one exposed to a factor postulated to cause a disease and the other unexposed. The groups are followed over time and their change in disease status is recorded during the study period. A retrospective cohort study starts when all of the disease cases have been identified. The history of each study participant is carefully evaluated for evidence of exposure to the agent under investigation.

Advantages: Because subjects are monitored over time for disease occurrence, cohort studies provide estimates of the absolute incidence of disease in exposed and non-exposed individuals. By design, exposure status is recorded before disease has been identified. In most cases, this provides unambiguous information about whether exposure preceded disease. Cohort studies are well-suited for studying rare exposures. This is because the relative number of exposed and

non-exposed persons in the study need not necessarily reflect true exposure prevalence in the population at large.

Disadvantages: Prospective cohort studies require a long follow-up period. In the case of rare diseases large groups are necessary. Losses to follow-up can become an important problem. Often quite expensive to run.

To assess the possible carcinogenic effects of radio-frequency signals emitted by cellular telephones, Johansen et al. (2001) conducted a retrospective cohort study in Denmark. Two companies that operate cellular telephone networks provided names and addresses for all 522,914 of their clients for the period 1982 to 1995. The investigators matched these records to the Danish Central Population Register. After cleaning the data 420,095 cellular telephone subscribers remained and formed the exposed cohort. All other Danish citizens during the study years became the unexposed cohort. The list of exposed and unexposed individuals were then matched with the national cancer registry. The resulting data allowed calculation of cancer incidence rates. Overall, 3,391 cancers had occurred among cellular telephone subscribers, compared with 3,825 cases expected based on age, gender, and calendar-year distribution of their person time at risk.

Reference: Johansen C, Boise J, McLaughlin J, Olsen J (2001). Cellular telephones and cancer — a nationwide cohort study in Denmark. Journal of the National Cancer Institute, 93: 203 - 237.

Case-control studies

Say we're interested in investigating risk factors for a rare disease such as bladder cancer in dogs. Imagine we have access to a perfect data set where we have the medical records for every dog in the country and details about things these dogs have been exposed to in their first year of life. For a given exposure (e.g. access to benzidine) we can present the data in a 2×2 table format, as shown in Table 9.

Table 9: Hypothetical data from a study of bladder cancer in a population of dogs.

	Disease +	Disease -	Total
Benzidene +	60	188,940	189,000
Benzidene -	57	278,943	279,000
Total	117	467,883	468,000

In this hypothetical example the risk of bladder cancer is 32 per 100,000 in those dogs exposed to benzidene in their first year of life, compared with 20 per 100,000 in those not exposed. The risk of bladder cancer is 1.6 times greater in exposed dogs than non-exposed dogs.

Now think of the logistics involved in carrying out this study. We would have to enroll 468,000 dogs, ask detailed questions about their management during the first year of life, then follow them for an extended period (years) to work out which of them got bladder cancer. A formidable task. A case-control design is intended to provide the same answer in a much simpler way by studying *all* of the dogs who got bladder cancer and a *sample* of dogs who did not.

Suppose we used a case-control approach where we investigate all 117 dogs with bladder cancer and a sample of controls chosen by selecting one out of every 1000 dogs who remained free of disease. The results are shown in Table 10. If we only had access to the information provided in Table 10 we wouldn't be able to calculate the risk of bladder cancer in either exposed or unexposed dogs because we don't know the size of the population at risk. What we can do

Table 10: Hypothetical data from a study of bladder cancer in a population of dogs. The data is comprised of 117 cases of bladder cancer and 468 controls selected at random from the population.

	Disease +	Disease -	Total
Benzidene +	60	189	249
Benzidene -	57	279	336
Total	117	468	585

however is work out the odds of cancer in the exposed and unexposed groups and compare them. The odds of cancer in benzidene exposed dogs is $60 \div 189 = 0.32$. The odds of cancer in non-exposed dogs is $57 \div 279 = 0.20$. The ratio of theses two odds is $0.32 \div 0.20 = 1.6$, which is the same as the risk ratio calculated earlier. The reason we got the same result should be obvious: $60 \div 189$ divided by $57 \div 279$ is the same as $60 \div 189{,}000$ divided by $57 \div 279{,}000$. It is the same ratio as before — the two denominators have simply been divided by 1000 as we have sampled only one out of every 1000 dogs who did not get bladder cancer.

This example demonstrates the usefulness of the case-control study design. Details collected on all identified cases and a selection of disease-negative animals ('controls') yields the same result as a very expensive (and usually impractical) study where every member of a population at risk is examined.

In a case-control study a group of cases and non-cases ('controls') are selected and we compare the frequency of exposure factors in the cases with that of the controls. Cases are those study subjects who have developed the outcome of interest whereas controls are those who have not developed the outcome of interest at the time of selection. The key thing is that the set of controls represent a set of individuals whose exposure to the factor of interest reflects the exposure in the population from which the cases were drawn. In most situations the individual is the unit of interest, this design applies equally as well to aggregates of individuals (such as litters, pens, and herds). Figure 13 is a diagramatic representation of the case-control design.

The key issue when designing a case-control study is to ensure that cases and controls are similar in every way except for the exposure factors hypothesised to be associated with the disease of interest. Controls should be drawn from the same general population as the cases — this is necessary to protect against the possible distortions from effect modifiers (confounders). There are three approaches that might be used to ensure that cases and controls are similar:

- Restricted sampling. If breed is a likely confounder you might only select one breed in the study (the dominant breed in the source population).
- Matching. Each case is matched with a control that has identical (or at least similar) values of the confounding variable (e.g. age and sex). This method provides direct control over known confounders and under certain conditions the efficiency of the analysis is improved. Disadvantages: (1) recruitment of suitable controls can be difficult (when it is difficult to find a suitable match); (2) you cannot quantify the effect of the matching variable on the risk of disease; (3) analysis of the data must take into account the effect of matching; (4) it is possible to overmatch, which decreases the efficiency of the study (and sometimes introduces bias).

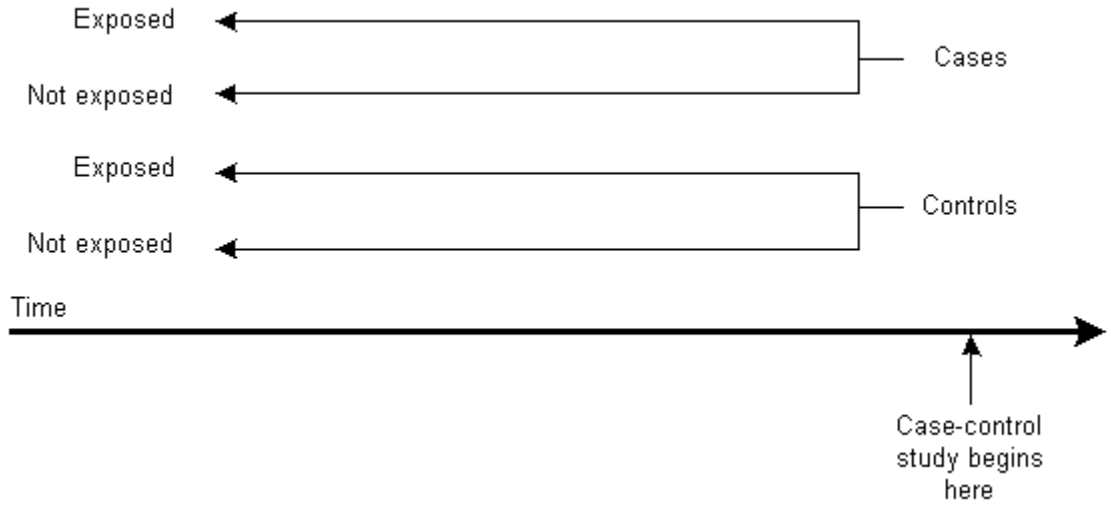


Figure 13: Schematic diagram of a case-control study.

- Analytical control. Multivariable regression techniques may be applied to remove the effect of known confounders.

Advantages: Case-control studies are an efficient method for studying rare diseases. Because subjects have experienced the outcome of interest at the start of the study, case-control studies are quick to run and are considerably cheaper than other study types.

Disadvantages: Case-control studies cannot provide information on the disease incidence in the studied population. The study is reliant on the quality of past records or recollection of study participants. It can also be very difficult to ensure an unbiased selection of the control group and, as a result, the representativeness of the sample selection process is difficult to guarantee.

Muscat et al. (2000) sought to test the hypothesis that cellular telephone use affects the risk of brain cancer. From 1994 to 1998 at five academic medical centres in the USA they recruited 469 cases aged 18 to 80 years with newly diagnosed cancer originating in the brain. Controls (n = 422) were inpatients without brain cancer at those hospitals, excluding those with leukaemia or lymphoma. Controls were sampled to match the cases on age, sex, race and month of admission. Each case and control was then interviewed about any past subscription to a cellular telephone service. Overall 14.1% of cases and 18.0% of controls reported ever having had a subscription for a cellular telephone service. After adjusting for age, sex, race, education, study centre, and month and year of interview, the risk of developing brain cancer in a cellular telephone user was estimated to be 0.85 (95% CI 0.6 – 1.2) times as great as in a non-user.

Reference: Muscat JE, Malkin MG, Thompson S, Shore RE, Stellman SD, McRee D (2000). Handheld cellular telephone use and risk of brain cancer. Journal of the American Medical Association, 284: 3001 - 3007.

Hybrid study designs

A **nested case-control study** is similar to a cohort study with the key difference that a sample of non-cases are selected for analysis (rather than the entire cohort, as in the case of a cohort study). Figure 14 shows a diagram of a nested case-control design.

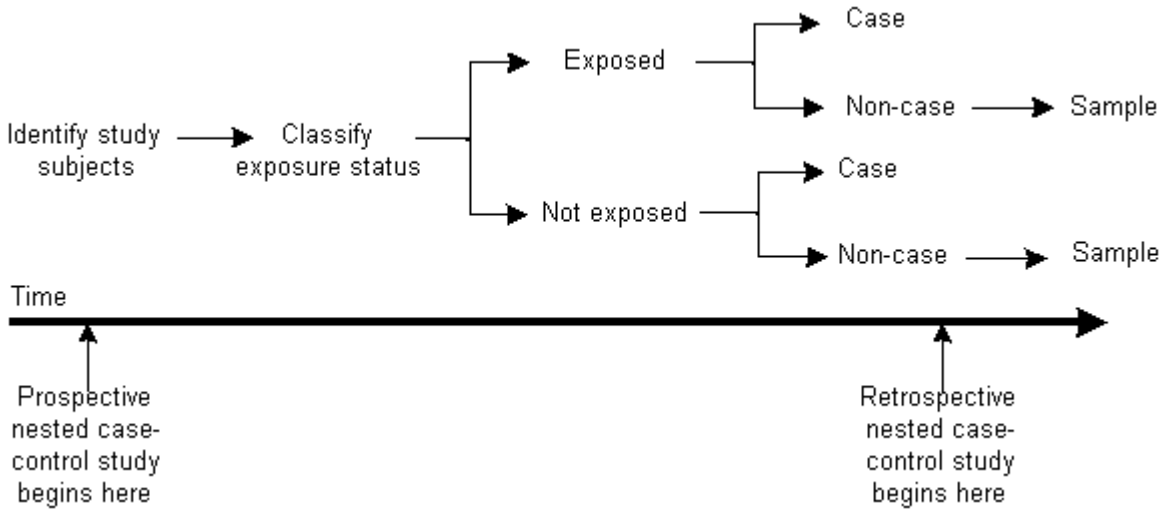


Figure 14: Schematic diagram of a nested case-control study.

Advantages: Nested case-control studies are useful when it is either too costly or not feasible to perform additional analyses on an entire cohort (e.g. if collection of specimens and laboratory analysis of specimens is expensive). Compared with standard case-control studies, nested studies: (1) can utilise exposure and confounder data originally collected before the onset of the disease, thus reducing potential recall bias and temporal ambiguity, and (2) include cases and controls drawn from the same cohort, decreasing the likelihood of selection bias. The nested case-control study is thus considered a strong observational study, comparable to its parent cohort study.

Disadvantages: A concern, usually minor, is that the remaining non-diseased persons from whom the controls are selected when it is decided to do the nested study, may not be fully representative of the original cohort due to death or losses to follow-up.

To determine if Helicobacter pylori infection was associated with the development of gastric cancer, Parsonnet et al. (1991) identified a cohort of 128,992 persons who had been followed since the mid-1960s. Of the original cohort, 189 patients developed gastric cancer. The investigators carried out a nested case-control study by selecting all of the 189 gastric cancer patients as cases and another 189 cancer-free individuals from the same cohort as controls. H. pylori infection status was determined using serum obtained at the beginning of the follow-up period. All total of 84% of the confirmed gastric cancer cases had been infected previously with H. pylori, while only 61% of the controls had been infected. This indicated a positive association between H. pylori infection and gastric cancer risk.

Reference: Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, Sibley RK (1991). Helicobacter pylori infection and the risk of gastric-carcinoma. New England Journal of Medicine, 325(16): 1127 - 1131.

In a **case-crossover study** a set of cases (subjects) is identified and a period of time before the onset of disease is selected (termed the case window) wherein the exposure to the risk factor of interest is evaluated. For each subject a second, non-overlapping time window (the control window) of the same length as the case window is selected, during which the subject did not experience the disease. This design is suitable for studying the transient effects of exposures that can vary over time and precipitate acute events (e.g. epilepsy episodes, asthma attacks). The design is efficient in that each case acts as its own control.

A study was conducted to determine if sleep disturbance was a risk factor for injury in children (Valent et al. 2001). A set of cases were identified and each child was asked if their sleep was disturbed in the 24 hours before the injury occurred (the case window) and in the 24 hours before that (the control window). Among 181 boys, 40 had less than 10 hours sleep on both days; 111 had less than 10 hours on neither day; 21 had less than 10 hours only on the day before the injury; and 9 had less than 10 hours sleep on the penultimate day before the injury. The odds ratio for injury, comparing days without and with 10 hours or more sleep was 2.33 (95% CI 1.02 – 5.79).

Reference: Valent F, Brusaferro S, Barbone F (2001). A case-crossover study of sleep and childhood injury. Pediatrics, 107: E23.

A **panel study** combines the features of cross-sectional and a prospective cohort designs. It can be viewed as a series of cross-sectional studies conducted on the same subjects (the panel) at successive time intervals (sometimes referred to as waves). This design allows investigators to relate changes in one variable to changes in other variables over time.

A **repeated survey** is a series of cross-sectional studies performed over time on the same study population, but each is sampled independently. Whereas panel studies follow the same individuals from survey to survey, repeated surveys follow the same study population (which may differ in composition from one survey to the next). Repeated surveys are useful for identifying overall trends in health status over time.

4.3 Experimental studies

Randomised clinical trials

The randomised clinical trial is the epidemiologic design that most closely resembles a laboratory experiment. The major objective is to test the possible effect of a therapeutic or preventive intervention. The design's key feature is that a formal chance mechanism is used to assign participants to either the treatment or control group. Subjects are then followed over time to measure one or more outcomes, such as the occurrence of disease. All things being equal, results from randomised trials offer a more solid basis for inference of cause and effect than results obtained from any other study design.

Advantages: Randomisation generally provides excellent control over confounding, even by factors that may be hard to measure or that may be unknown to the investigator.

Disadvantages: For many exposures it may not be ethical or feasible to conduct a clinical trial (e.g. exposure to pollution). Expensive. Impractical if long periods of follow-up required.

Bacterial vaginosis affects an estimated 800,000 pregnant women each year in the USA and has been found to be associated with premature birth and other pregnancy complications. To determine whether treatment with antibiotics could reduce the incidence of adverse pregnancy outcomes, Carey et al. (2000) screened 29,625 pregnant women to identify 1953 who had bacterial vaginosis, met certain other eligibility criteria, and consented to participate. Women were randomly assigned to receive either: (1) two 2 gram doses of metronidazole, or (2) two doses of a similar-appearing placebo.

Bacterial vaginosis resolved in 78% of women in the treatment group, but in only 37% of women in the placebo group. Pre-term labour, postpartum infections in the mother or infant, and admission to the neonatal intensive care unit were equally common in both groups.

Reference: Carey JC, Klebanoff MA, Hauth JC, Hillier SL, Thom EA, Ernest JM et al. (2000). Metronidazole to prevent preterm delivery in pregnant women with asymptomatic bacterial vaginosis. New England Journal of Medicine, 342: 534 - 540.

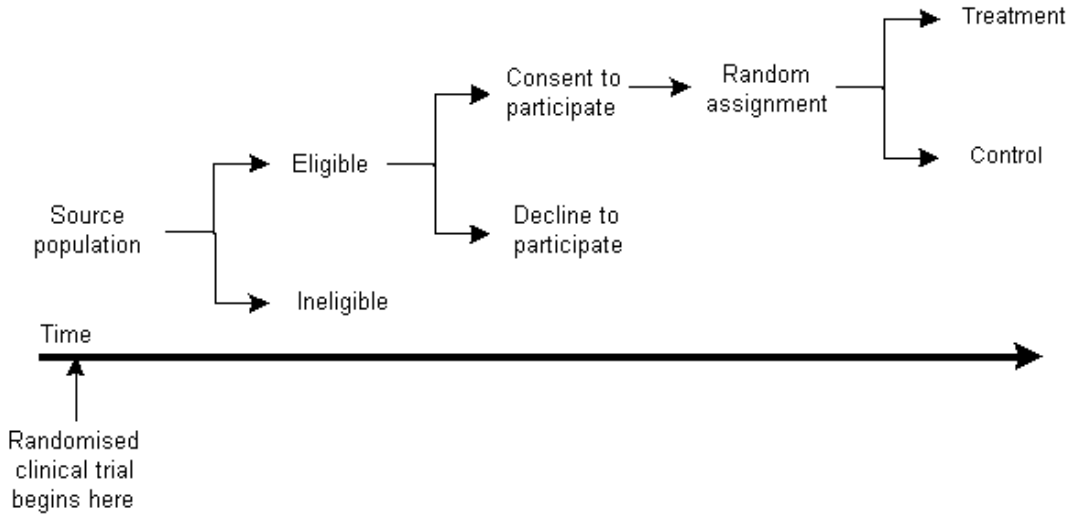


Figure 15: Schematic diagram of a randomised clinical trial.

Community trials

Instead of randomly assigning individuals to treatment or control groups, community trials assign interventions to entire groups of individuals. In the simplest situation one group (community) receives the treatment and another serves as a control.

4.4 Comparison of major the major study designs

Cohort studies involve enumeration of the denominator of the disease measure (individual time at risk) while case-control studies only sample from the denominator. Cohort studies therefore provide an estimate of incidence and risk whereas case-control studies can only estimate ratios. Prospective cohort studies provide the best evidence for the presence of cause-effect relationships, because any putative cause has to be present before disease occurs. Since these study designs are based on observation within a largely uncontrolled environment it is possible that there are still other unmeasured factors which produce cause-effect relationships that might be identified. The prospective cohort study is inefficient for studying rare diseases, which is a particular strength of the case-control study. A carefully designed cross-sectional study is more likely to be representative of the population than a case-control study.

Table 11: Comparison of the features of the cohort, case-control and cross-sectional study designs.

Criteria	Cohort	Case-control	Cross-sectional
Sampling	Separate samples of exposed and non-exposed individuals	Separate samples of diseased and non-diseased individuals	Random sample of study group
Time	Usually prospective (but may be retrospective)	Usually retrospective	Single point
Causality	Causality through evidence of temporality	Preliminary causal hypothesis	Association between disease and risk factor
Risk	Incidence risk, incidence rate	None	Prevalence
Measure of assocation	Risk ratio, odds ratio	Odds ratio	Risk ratio, odds ratio

5 Error in epidemiological research

By the end of this unit you should be able to:

- Explain the difference between random error and bias and how each can affect the results of epidemiologic research.
- Describe the key features of selection and misclassification bias. Provide examples of selection and misclassification bias. Explain how you might minimise each type of bias when conducting an observational study.
- Describe the common sources of bias in each of the major epidemiological study designs (ecological, cross-sectional, cohort, and case-control studies).
- Explain the difference between confounding and interaction, with examples.

5.1 Sources of error

When you derive an estimate from a population from a sample you want it to be precise and accurate. A **precise** estimate has confidence intervals that are small. An **accurate** estimate has confidence intervals that are centred on the true population value. There are two types of error that can occur in epidemiologic research: random error and bias. The difference between random error and bias is explained in Figure 16.

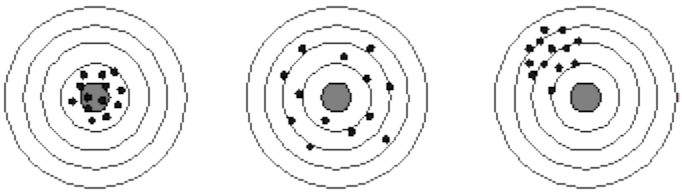


Figure 16: The distribution of bullets fired at the target on the left show little evidence of random error and bias. The distribution of the bullets fired at the centre target show a high degree of random error and a low degree of bias. The distribution of the bullets fired at the target on the right show a low degree of random error and a high degree of bias.

5.2 Random error

Random error occurs by chance. A random selection of individuals taken to make up a sample will differ slightly from each other. These differences will result in sample estimates that differ slightly from each other and also from the target population. Random error is the inherent error that arises from using a sample to make a measurement of a population. The influence of random error may be reduced by:

1. Increasing the size of the sample taken. Using the central limit theorem it can be demonstrated that a fourfold increase in sample size will result in a halving of the confidence interval.

2. Modifying the sample selection procedure to ensure that only the target group is sampled. For example, you may be interested in the performance of only one particular breed of dairy cow. You can design the study to ensure that you sample animals only from farms that contain this breed of cow. Stratified sampling is a technique that reduces sample variance

by dividing the population into individual strata. Each stratum contains individuals that are similar, and so the variance within strata is less than the variation between strata. You would typically obtain samples from individual strata that have less variation than similar-sized samples obtained from the whole (unstratified) population.

3. Using an appropriate scale of measurement. Ratio estimators may result in a reduction in confidence intervals in some situations. Suppose, for example, you wish to determine whether farmed lambs have reached the correct weight for sale. You could take a sample of lambs and estimate the average weight of the sample (and an associated confidence interval). If the weight of lambs in the population is variable and you do not select a large sample it is likely that the associated confidence interval will be wide (and will include the target value). An alternative is to dichotomously classify each lambs weight within the sample with respect to the target weight (i.e. describe it as either above or below target weight). You can then calculate an estimate of the proportion of lambs that have obtained target weight (along with associated confidence intervals). You are more likely to produce narrow confidence intervals for this ratio estimate and are thus able to make a more confident decision regarding the sale of the lambs.

5.3 Bias

Bias is caused by systematic error. A systematic error is one that is inherent to the measurement technique being used. It results in a predictable and repeatable error for each observation. Bias results in observed effect estimates which differ from those which truly exist in the target population (other than differences due to random error). There are two broad types of bias: **selection bias** and **misclassification bias**. Some authors consider confounding to be a type of bias. In this discussion, we consider it as a distinct entity as a source of error in epidemiological research.

Selection bias

Selection bias is caused by the procedures used to select units that are included in a study. Selection bias occurs when these procedures results in the observed measure of effect that is different for study participants and study non-participants. The different types of selection bias include:

- Surveillance bias: if disease is asymptomatic or mild, it is more likely to be detected in persons under frequent medical surveillance.
- Referral bias: differential referral patterns are a source of bias in hospital-based case-control studies.
- Non-response bias: non-response or refusal to participate in a study.
- Length of stay bias: for hospital-based case-control studies, cases and controls should ideally be selected by a scheme that is equivalent to sampling admission logs (incident cases) rather than the hospital register of current patients (prevalent cases).

- Survival bias: the introduction of insulin has increased the lifespan of diabetic patients, producing an apparent increase in the prevalence of the disease.

Selection bias is usually cited as an issue for observational studies where associations between exposure and disease are being investigated. However, in descriptive studies (for example, where the frequency of disease is being described for a given population), the term selection bias can also be used to describe the situation when the frequency of disease in the group studied is not representative of that in the target population. This might occur when the sampling frame is not representative of the target population, and/or when response rates and/or withdrawal rates are high. Note that, in descriptive studies, the aim is to estimate disease frequency rather than to investigate associations between exposure and disease. So, in this case, selection bias will result in disease frequency estimates that differ from that in the target population. Selection bias cannot be controlled for using analytical techniques. The basic options for avoiding selection bias are as follows:

- Ensure that study participants are selected at random from the eligible population. Random selection does not ensure that the study population is representative of the eligible population. Rather, random sampling allows the probability of differences between the study population and the eligible population to be assessed. The probability of substantial differences between the two is greatest when sample sizes are small.
- Ensuring that response rates are high amongst the study population.
- Ensuring that withdrawal rates are low amongst the study population.
- In observational studies, consider the 'forces' which result in individuals being selected.

Misclassification bias

Misclassification (information) bias is due to errors in the information that is recorded for study participants. The different types of misclassification bias include:

- Recall bias: cases are generally better at recalling past exposure events, compared with non-cases.
- Interviewer bias: a potential problem when interviewers are privy to the hypothesis under investigation.
- Prevarification bias: subjects in a study may have ulterior motives for deliberately overestimating exposure to a hypothesised causal agent (e.g. compensation claims).
- Improper analysis bias: if one matches cases and controls on a variable that is associated with the study exposure, an analysis that ignores the matching will yield a disease exposure odds ratio that is biased towards unity.

Misclassification bias can be differential or non-differential. Errors on one axis (e.g. exposure) that are independent of the other axis (disease) result in non-differential misclassification.

Non-differential misclassification can be random or systematic. Non-differential random misclassification occurs when there are random errors in the instrument measuring exposure. Non-differential systematic misclassification occurs when the presence of exposure is systematically over- or under-estimated for both case and control groups. Both non-differential random and non-differential systematic misclassification results in the observed measure of association being biased towards the null.

When the measurement error and resulting misclassification occur to a greater extent in one group than another they are described as being differential. The effects of differential misclassification are generally harder to predict than those of non-differential misclassification. Differential random misclassification occurs when there are random errors in the instrument measuring exposure in either the case or control group. Differential systematic misclassification occurs when there are systematic errors in the instrument measuring exposure in either the case or control group. Both differential random and differential systematic misclassification have unpredictable effects on the observed measure of association. This type of misclassification can be very difficult to deal with because, unless you have some idea of how much is occurring and where it is occurring, you cannot estimate the magnitude of its effect or the direction of its effect (i.e. whether it shifts the observed measure of association away from, or towards the null).

Note that the foregoing discussion applies only to studies that ascertain exposure and disease status for individual subjects within a population. In ecological studies, in which the effect of an exposure is estimated by correlating disease rates across groups with differences in their exposure prevalence, non-differential misclassification of exposure can actually lead to an inflated estimate of the influence of exposure on disease risk. See Brenner et al. (1992) for further discussion of this topic.

Misclassification bias cannot be controlled for using analytical techniques. Options for avoiding misclassification bias are as follows:

- Ensure that exposure and disease status are assessed independently — i.e. assess one without knowledge of the other.
- Use a rigorous and biologically valid method for determining the presence of disease and exposure.
- Use complete and detailed sources of information (i.e. complete exposure histories).
- Use objective measures where available (e.g. liveweights, fleece weights, milk production records, pregnancy testing results, laboratory measurements).

We usually consider that laboratory measurements apply most frequently to classification of disease status. However, laboratory measures may also be useful for determining exposure status. For example, if the exposure of interest was contamination with a chemical, laboratory estimates of residues may be useful. Trace element, macromineral and parasitological status of livestock may be assessed using laboratory tests when we are interested in the effects of these exposures rather than the determinants of these conditions.

5.4 Confounding and interaction

A confounder is an extraneous factor that wholly or partially accounts for the observed effect of a risk factor on disease. Interaction, as distinct from confounding, is the interdependent operation of two or more factors to produce an unanticipated effect.

Confounding

The association between an exposure and a given outcome may be strengthened, weakened, or eliminated by a confounding variables. The following criteria are used to determine whether or not a variable confounds the relationship between an exposure and an outcome:

1. The confounding variable is causally associated with the outcome; and
2. The confounding variable is noncausally associated with the exposure; and
3. The confounding variable and the exposure variable are on two separate causal pathways to the outcome.

Figure 17 shows the relationship between three factors: exposure, confounder, and outcome. Think of the arrows that connect each of the three factors as water pipes. The direction of the arrow (which represents the association between one variable and another) can be thought of in terms of a flow of water and the strength of the association can be thought of in terms of the water pressure travelling through the pipe. Our interest is to determine the water pressure in the pipe linking exposure and outcome (i.e. the strength of the association between exposure and outcome). When there is no confounding, water arising from the exposure arrives at the outcome directly (at a given pressure). When confounding is present, water arising from exposure arrives at the outcome from two sources: directly from the exposure and via the confounder. In this case, the presence of the confounder changes the strength of association between the exposure and outcome (changing the water pressure in the pipe from connecting the exposure with the outcome).

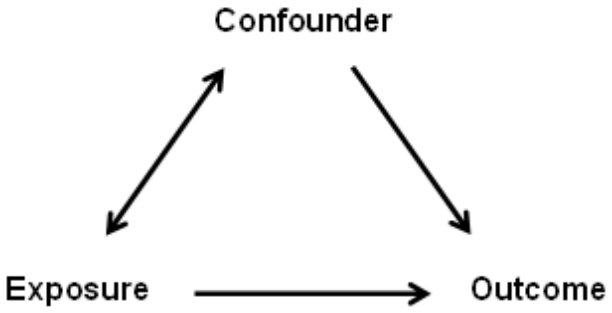


Figure 17: Schematic diagram of the relationship between an exposure, outcome, and confounder. The confounder is causally associated with the outcome of interst and either causally or noncausally associated with the exposure. In the above diagram a unidirectional arrow indicates the association is causal; a bidirectional arrow indicates a noncausal association.

Does drinking confound the association between smoking (the exposure) and laryngeal cancer (the outcome)?

1. Is drinking causally associated with laryngeal cancer? Yes.
2. Is drinking noncausally associated with smoking? Yes.
3. Is the association between drinking and laryngeal cancer and the association between smoking and laryngeal cancer on two separate causal pathways? Yes.

We conclude that drinking confounds the association between smoking and laryngeal cancer.

Once we have determined that a variable is likely to be a confounder, we need to consider the likely direction of its effect. Does the presence of the confounder strengthen or weaken the observed association? To answer this question, take the following four-step approach:

Step 1. Construct a 2 × 2 table with the two levels of exposure as rows and the two levels of disease status as columns. Figure 18 shows the 2 × 2 table for the smoking and laryngeal cancer example.

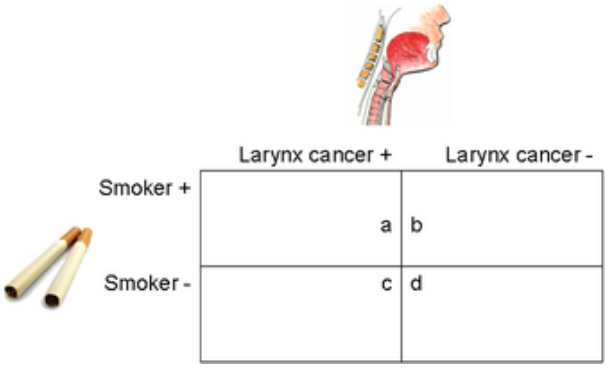


Figure 18: 2 × 2 table with smoking (rows) as the exposure and laryngeal cancer (columns) as the outcome.

Step 2. Think about the effect of the confounder on each level of exposure. In this example, we ask the question: who are bigger drinkers — smokers or non-smokers? Answer: drinking is likely to be positively associated with smoking. Indicate this effect on your 2 × 2 table, as shown in Figure 19.

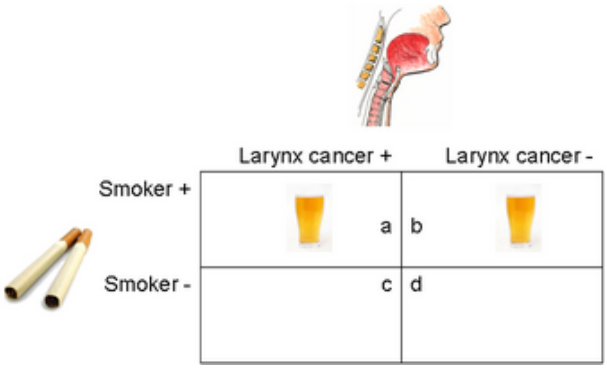


Figure 19: 2 × 2 table for the association between smoking and laryngeal cancer showing the likely effect of drinking on smoking.

Step 3. Think about the effect of the confounder on each level of the outcome. Is drinking likely to be positively or negatively associated with laryngeal cancer? Answer: drinking is likely to be positively associated with laryngeal cancer. Again, indicate this effect on your 2 × 2 table, as shown in Figure 20.

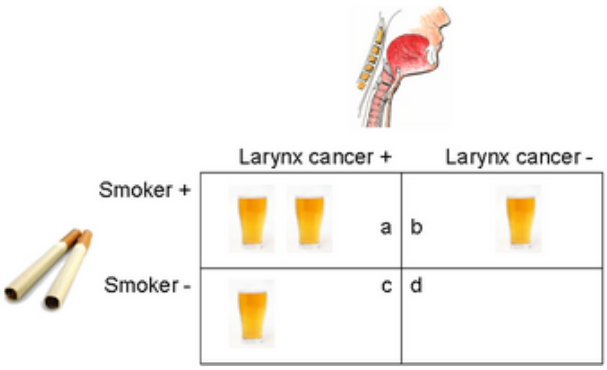


Figure 20: 2 × 2 table for the association between smoking and laryngeal cancer showing the likely effect of drinking on laryngeal cancer.

Step 4. Interprete the results. Drinking is positively associated with smoking, so we anticipate an overrepresentation of subjects in the exposure-positive (a and b) cells of the 2 × 2 table. Drinking is positively associated with laryngeal cancer, so we anticipate an overrepresentation of subjects in the outcome-positive (a and c) cells of the 2 × 2 table. Based on these judgements, the size of the a cell will be exaggerated. This means that the proportion of smokers with laryngeal cancer will be increased, resulting in a strengthening of the primary exposure-outcome association.

We conclude that drinking confounds the association between smoking and laryngeal cancer, making the association between smoking and laryngeal cancer stronger than what it actually is.

Interaction

In addition to confounding, extraneous factors can influence the strength of association between exposure and outcome. This biological phenomenon is known as interaction (also known as effect modification). Interaction refers to a difference in the effect of one exposure according to the level of another. Like confounding, interaction is due to the influence of an extraneous factor.

An example of interaction occurred in a cohort study of elderly people conducted by Fransen et al. (2002). In this study the chance of death or institutionalisation within 2 years was much higher for those who had previously suffered a hip fracture at the start of the study. The excess risk associated with hip fracture was significantly higher for men than women. This is an example of interaction between hip fracture status at study start (yes or no) and sex.

Reference: Fransen M, Woodward M, Norton R, Robinson E, Butler J, Campbell A (2002). Excess mortality or institutionalisation following hip fracture: men are at greater risk than women. Journal of the American Geriatrics Society, 50: 685 - 690.

There are three types of interaction:

- **Unilateralism**: where A has no effect in the absence of B, but a considerable effect in the presence of B.

- **Synergism**: where the effect of A is in the same direction, but stronger in the presence of B.
- **Antagonism**: where the effect of A works in the opposite direction when acting in the presence of B, to the direction in which it acts in the absence of B.

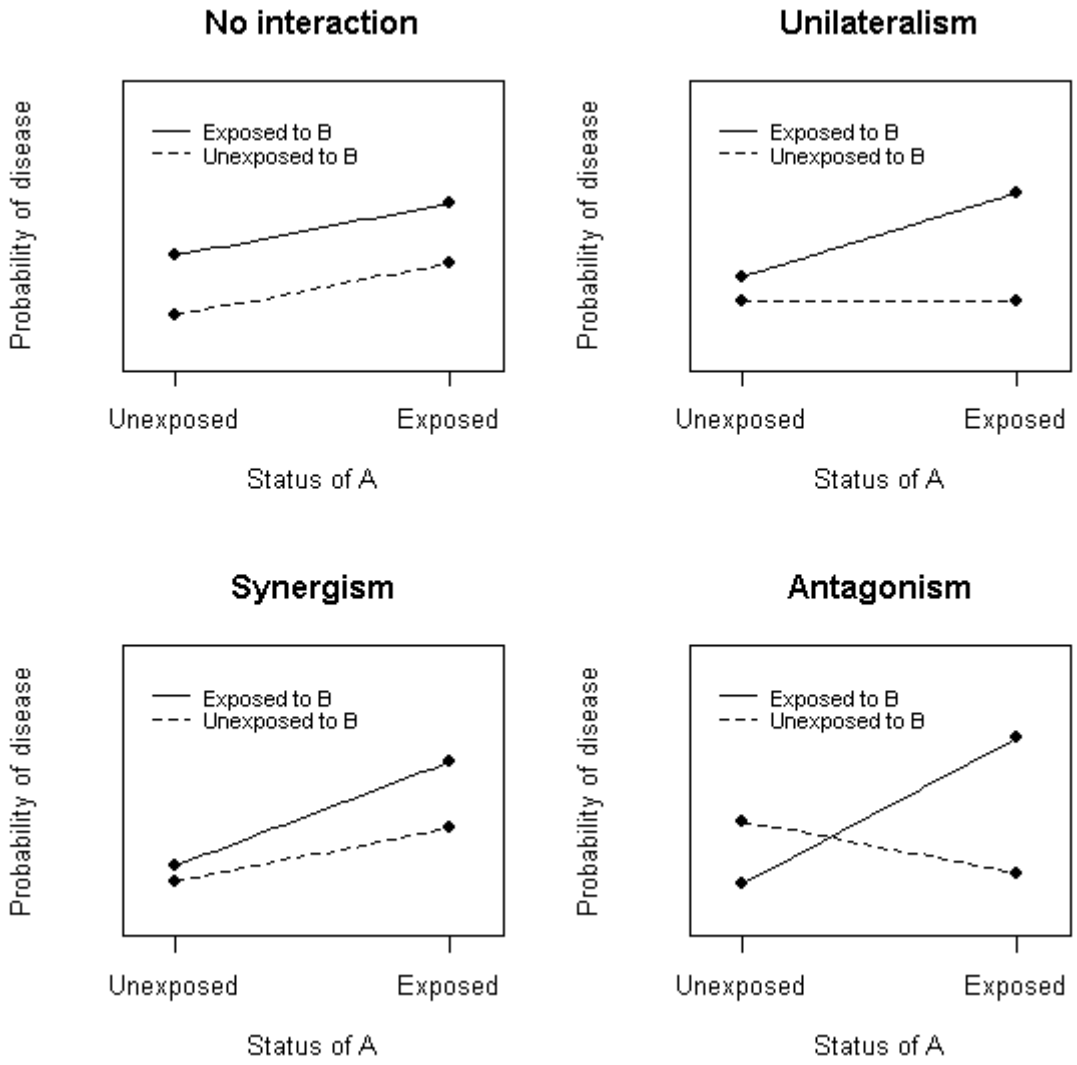


Figure 21: Interaction diagrams showing the four types of interaction that may occur between risk factors A and B.

Identifying the presence of confounding and interaction

Confounding and interaction are different phenomena. A variable may manifest itself as either a confounder or interactive factor, as neither, or both. Different strategies are used to identify the presence of confounding and interaction in an exposure and outcome data set. In the first instance, data are stratified to calculate strata-level measures of association. If measures

of association differ among strata significantly, then we conclude that interaction is present. If strata-level measures of association are not significantly different, interaction is said to be absent and the presence of confounding is investigated.

To be quantitative about identifying the presence of confounding we compare the crude measure of association with an adjusted measure (e.g. the Mantel-Haenszel adjusted odds ratio). If there are i strata and the total number of subjects in each strata is T_i, the crude odds ratio equals:

$$OR_{\text{crude}} = \frac{\sum_i a_i \sum_i d_i}{\sum_i b_i \sum_i c_i} \tag{15}$$

and the Mantel-Haenszel adjusted odds ratio equals:

$$OR_{\text{M-H}} = \frac{\sum_i \frac{a_i d_i}{T_i}}{\sum_i \frac{b_i c_i}{T_i}} \tag{16}$$

The formula for the adjusted odds ratio is provided here simply to give you an idea of what the adjustment process involves. Formulae for adjusting other measures of association are provided in many standard epidemiological texts — Elwood (1998) provides as very clear description of the approach.

If the adjustment changes the interpretation of the exposure and outcome relationship, we have quantitative evidence to support the notion that confounding is present. If the adjustment does not change the interpretation of the exposure-outome relationship, we conclude that the effect of confounding is either small or absent. A summary of the approach for distinguishing confounding and interaction in a data set is shown in Figure 22.

Do not confuse confounding and interaction. Interaction implies different risk/odds ratios across different strata. Because the strata-level measures of association differ, summary measures are not appropriate. Confounding implies that strata will have risk/odds ratios that are similar to each other and that the summary (adjusted) risk/odds ratio differs from the crude (unadjusted) risk/odds ratio.

Methods for dealing with confounding

Restriction

Think about the association between the number of children a woman has had and the risk of being diagnosed with breast cancer and the confounding effect of age. To deal with age as a confounder we could only include those subjects that were of a certain age in the study population. We could do this with either a cohort or a case-control design. Restriction is clearly an effective method, as it leaves no possibility of confounding, but obviously the disadvantage is that the study then becomes specific to a particular age group, and we cannot generalise the study beyond that target population.

Randomisation

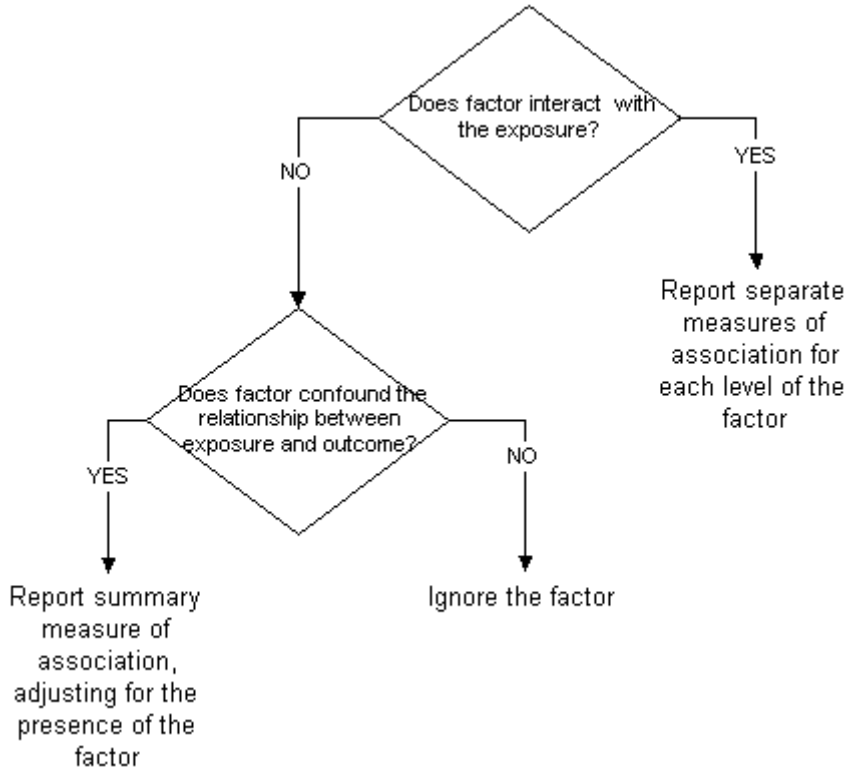


Figure 22: Strategy for distinguishing confounding and interaction.

Randomisation is an option for dealing with confounding in prospective intervention studies assessing the effects of an ethical, practical and acceptable intervention which is thought to be beneficial and not likely to be harmful. Randomisation cannot be applied to retrospective studies.

The principle of randomisation is that from a pool of study participants, subjects are randomly assigned to exposure and non-exposure groups. The definition of random is such that each subject in the study has the same chance of being allocated to a particular group, and that the chance of one individual being allocated to one group is not influenced by the allocation of any other member to the group. The advantage of randomisation is that, given large sample sizes, it is likely to produce groups which are similar even in respect to variables which have not been anticipated, designed, or measured.

Stratification

The best way to adjust for a single confounder is to examine exposure-outcome relationships within levels of the confounder. Within each of the confounder levels, there will be no confounding because exposed and non-exposed subjects will all have the same level of the confounder. If the size of the exposure-outcome association is the same at all levels (or strata) of the confounder, then statistical methods can be used to combine the stratum-specific estimates of effect to give an estimate of effect that is adjusted for the confounder. Note that the term effect modification is used to describe the situation where the exposure-outcome relationship varies according to the level of the confounder.

Matching

Matching each exposed subject to an unexposed subject with the same level of a confounder will reduce selection bias. For example, in the hypothetical smoking/heart disease study, smokers and non-smokers could be matched according to sex. When a male smoker is recruited into the study, he is matched with a male non-smoker. When a female smoker is recruited she is matched to a female non-smoker. This will obviously lead to identical percentages of men and women among smokers and non-smokers.

Matching of exposed and non-exposed subjects is only possible in studies where subjects are recruited on the basis of their exposure status. It is not possible in case-control studies, where subjects are recruited according to their outcome status (presence or absence of outcome). Thus, in a case-control study of smoking and coronary heart disease, people with heart disease could be matched by sex to people without heart disease. This would result in cases and controls having the same percentage of male and females but would not lead to an even distribution of sex among smokers and non-smokers. The latter condition is the important one for control of confounding by sex. Matching is an excellent design strategy for control of confounders in cohort studies. However, it is inappropriate for this purpose in case-control studies.

The purpose of matching in case-control studies is to improve the statistical power of the study. If matching is done to improve power in a case-control study, then the data analysis should take this matching into account. Otherwise, bias can be introduced into the study.

Multivariate methods

Whereas stratification is an excellent method for controlling a single confounder, multivariate methods (statistical modeling) is required if there are multiple confounders. One disadvantage of modeling as a means to control confounding is that the investigator is distanced from the mechanics of the data analysis: stratification permits a much better 'feel' for the data and should always precede modeling.

A worked example

Siscovick et al. (1984) conducted a case-control study to evaluate the relationship between primary cardiac arrest and habitual vigorous exercise. They reported the data shown in Table 12:

Table 12: Results of a case-control study evaluating the relationship between habitual exercise and primary cardiac arrest (Siscovick et al. 1984).

Non-smokers	Diseased	Non-diseased	Total
Exercise -	36	24	60
Exercise +	32	70	102
Total	68	94	162

We would like to know: (1) if there is an interaction between smoking and vigorous habitual exercise on the risk of primary cardiac arrest, and (2) if smoking confounds the association between vigorous habitual exercise and primary cardiac arrest. First, we check for evidence of interaction:

Smokers	Diseased	Non-diseased	Total
Exercise -	40	17	57
Exercise +	25	22	47
Total	65	39	104

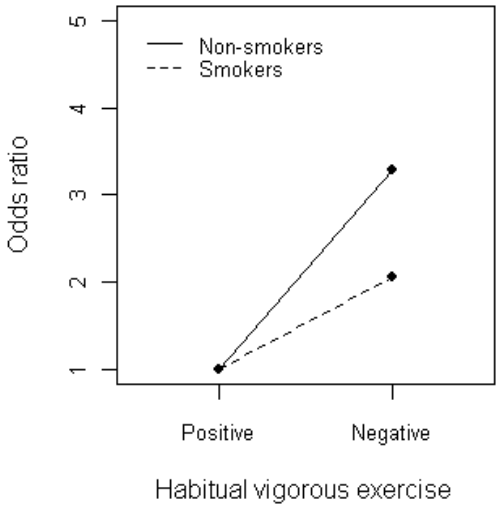


Figure 23: Interaction plot showing the nature of the interaction between smoking and habitual vigorous exercise on the risk of primary cardiac arrest.

The odds of primary cardiac arrest for non-smokers that did not undertake habitual vigorous exercise was 3.28 (95% CI 1.69 – 6.38) times that of those who did exercise habitually. The odds of primary cardiac arrest for smokers that did not undertake habitual vigorous exercise was 2.07 (95% CI 0.92 – 4.64) times that of those who did exercise habitually. There appears to be a synergistic interaction between smoking, exercise, and risk of primary cardiac arrest (Figure 23).

Although we have evidence to suggest the presence of interaction, we need to test the hypothesis that the strata-level odds ratios are the same using the chi-squared test of homogeneity. The test of homogeneity test statistic is compared with a chi-squared distribution with n - 1 degrees of freedom (where n is the number of strata). A test of homogeneity of the stratified odds ratios produces a χ^2 test statistic of 1.03. Since there are two strata the comparison has 1 degrees of freedom and the associated P-value is 0.31. We accept the null hypothesis and conclude that the stratum-specific odds ratios are the same (that is, there is no significant interaction). We now use the four criteria outlined above to assess whether smoking is a confounder in the relationship between habitual vigorous exercise and primary cardiac arrest.

Is smoking causally associated with primary cardiac arrest?

	Diseased	Non-diseased	Total
Smoking +	65	39	104
Smoking -	68	194	162
Total	133	133	266

The odds of primary cardiac arrest for smokers was 4.75 (95% CI 2.93 – 7.71) times that of non smokers. We have evidence that smoking is associated with cardiac arrest. A review of the relevant literature would also support the notion that this association is causal.

Is smoking noncausally associated with habitual exercise?

	Exercise+	Exercise -	Total
Smoking +	47	57	104
Smoking -	102	60	162
Total	149	117	266

The odds of being a habitual exercise for smokers was 0.49 (95% CI 0.29 – 0.80) times that of non smokers. It is reasonable to conclude that being a smoker is noncausally associated with (lack of) habitual exercise.

Is the link between habitual exercise and cardiac arrest and smoking and cardiac arrest on two separate causal pathways? Lack of habitual exercise and smoking increase the risk of cardiac arrest by two independent physiological mechanisms. It is reasonable to assume that they are on two separate causal pathways.

Does the strength of the association between habitual exercise and primary cardiac arrest change when you account for the presence of smoking? Compare the crude odds ratio with the Mantel-Haenszel adjusted odds ratio. The odds of primary cardiac arrest in those who undertook habitual vigorous exercise was 2.99 (95% CI 1.81 – 4.95) times greater than those who did not. Apply the Mantel-Haenszel procedure to produce an adjusted odds ratio. After adjusting for smoking status, the odds of primary cardiac arrest for those that did not undertake habitual vigorous exercise was 2.72 (95% CI 1.46 – 5.06). The ratio of the crude odds ratio to the adjusted odds ratio is $2.99 \div 2.72 = 1.10$. We conclude that smoking confounds the association between habitual vigorous exercise and risk of primary cardiac arrest (using a relative difference of greater than 10% to 15% between the crude and adjusted odds ratio as an objective indicator of the presence of confounding).

6 Causation

By the end of this unit you should be able to:

- Explain the difference between association and causation in epidemiological research.
- Define component, sufficient, and necessary causes (with examples).
- List and briefly explain each of Hill's criteria for causation.

6.1 Association versus causation

A fundamental objective of epidemiologic research is to identify the causes of disease through the study of the distribution of cases within groups of individuals with identified characteristics, such as different levels of exposure to some agent (e.g. exposure to a drug or chemical). Knowledge of what causes disease allows us to develop prevention strategies by targeting those risk factors influential in determining the likelihood that disease will occur.

Using this approach we need to be aware of the difference between **association** and **causation**. Association is a quantitative measure of the relationship between two or more factors or outcomes. A cause, on the other hand, is an event, condition, or characteristic without which disease cannot occur (Rothman 1976).

A study conducted in the 1980s found that dairy herds milked by staff who wore shorts and aprons during milking were more likely to be positive for leptospirosis. These findings lead to the question: do shorts and plastic aprons cause leptospirosis, or are they associated with the presence of leptospirosis? Obviously, shorts and aprons are associated with the presence of leptospirosis, and in this example their use was a marker (i.e. a proxy variable) for other causative factors, such as (for example) herd size.

Reference: Mackintosh C, Schollum L, Harris R, Blackmore D, Willis A, Cook N, Stoke J (1980). Epidemiology of leptospirosis in dairy farm workers in the Manawatu. Part I: A cross-sectional serological survey and associated occupational factors. New Zealand Veterinary Journal 28: 245 - 250.

When a previously unrecognised disease is identified epidemiological research usually starts with case reports and case series that describe the condition and provide evidence that the disease can occur repeatedly. Descriptive studies follow, where the distribution of disease is documented according to individual, place, and time. Descriptive studies are useful because they provide a rich source of hypotheses about factors that are associated with, or cause the disease. Analytical studies (i.e. cross-sectional, cohort, case-control studies) provide a means for testing the hypotheses generated from descriptive studies.

Because epidemiology is predominantly an observational (i.e. non-experimental) science that draws its data from the uncontrolled conditions, we need to be aware that bias, confounding, and chance may provide alternative explanations for the associations that we might identify. If these issues are thought to be present then further analytical studies need to be undertaken to account for them. Once we are confident of the validity of the associations that have been observed (i.e. we're confident bias, confounding, and chance are not present) attention turns towards establishing if the relationships between the identified risk factors and disease are causal (Figure 24 provides an outline of the process).

Whereas the identification of association is predominantly a quantitative process, identifying causal relationships is largely subjective and based on judgement. Over the years, several authors (e.g. Koch, Evans, and Hill) have defined criteria that might be used to help identify is an observed relationship is causal.

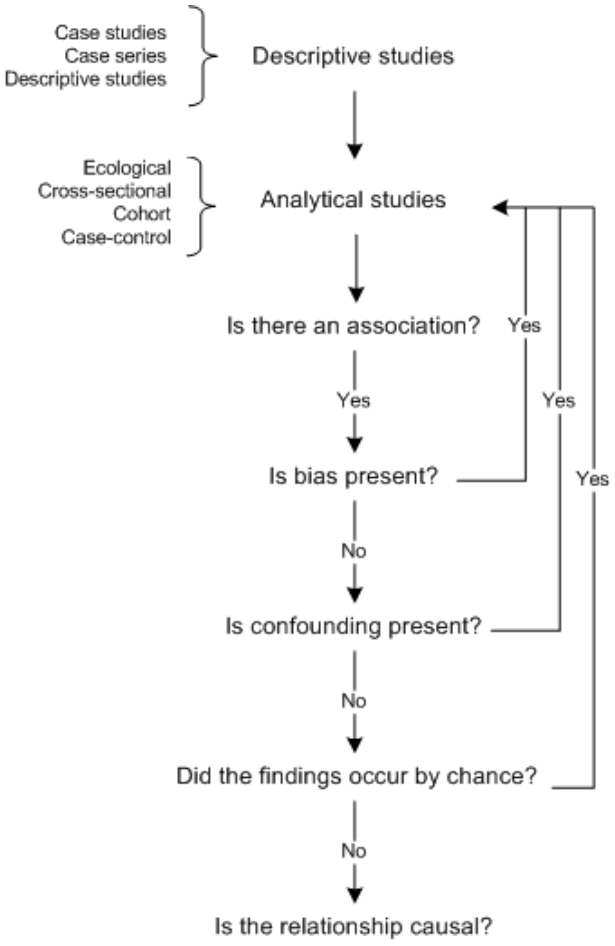


Figure 24: Flow diagram outlining the typical sequence of events in the epidemiological investigation of disease.

6.2 Component, sufficient, and necessary causes

Causes have the following characteristics: (1) they must precede the effect, (2) they can be either host or environmental factors (e.g. characteristics, conditions, actions of individuals, events, natural, social or economic phenomena), and (3) they can be either positive (the presence of an exposure) or negative (the absence of exposure, such as vaccination).

The key thing is that one or more causes are determinants of many of the diseases that we deal with. On one hand we might have a condition such as anthrax, which has a single cause (exposure to *Bacillus anthracis*). For other diseases, such as lameness in dairy cattle, there might be multiple causes (e.g. poor hoof condition, injury, age). This said, it is easiest to conceptualise causation by regarding causal factors as the pieces of a pie. Disease occurs when we have assembled enough causal factors to to produce a full pie. For some diseases (especially infectious conditions) it may be that exposure to the infectious agent will cause disease: in this situation there is only one piece to the pie. For other diseases there may be many reasons why some exposed individuals don't develop the disease yet others do: in this situation, the pie is made up of many pieces. The following terms are used when talking about causation:

- Component causes are conditions that are causally related to the presence of disease (the pieces of the pie). Factors such as high cholesterol, smoking, lack of exercise, genetics, and the presence of concurrent diseases are all component causes of coronary heart disease in humans.
- Sufficient causes are the set of conditions without any one of which disease would not have occurred (the whole pie). Sufficient causes are not usually a single factor but several. Accumulation of a set of sufficient causes is synonymous with occurrence (although not necessarily diagnosis) of disease.
- A necessary cause is one that must be present for the disease to occur (the most important piece of the pie). If chicken salad has been identified as sufficient causes of salmonellosis in a foodborne disease outbreak, *Salmonella spp.* would be a necessary cause of diarrhoea.

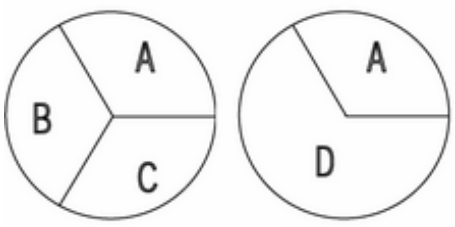


Figure 25: This diagram represents a disease that has two sufficient causal complexes: the first having three component causes and the second having two. A, B, C, and D are component causes. A is a necessary component cause.

Whereas infection with Mycobacterium tuberculosis is a necessary cause for tuberculosis disease, it is not a sufficient cause since many animals may harbor small foci of tuberculosis without developing tuberculosis.

Tobacco smoking is a sufficient cause of lung cancer, but so is exposure to radon or certain occupational chemicals in non smokers.

Coronary heart disease in humans has no necessary cause, but rather a range of component causes which become sufficient when some or all occur together in individuals at levels that cumulate and interact to result in disease.

Koch (1884) was the first to provide a framework for identifying causes of infectious disease. Koch specified that the following criteria (known as Koch's postulates) had to be met before an agent could be considered as the cause of a disease:

1. The agent has to be present in every case.
2. The agent has to be isolated from the affected individual and grown in pure culture.
3. The agent has to cause disease when inoculated into a susceptible animal and the agent must then be able to be recovered from that animal and identified.

In the late nineteenth century Koch's postulates brought a degree of order and discipline to the study of infectious diseases, although the key assumption of 'one-agent-one-disease' was highly restrictive since it failed to take account of diseases with multiple aetiologic factors, multiple effects of single causes, carrier states, and non-agent factors (such as age and sex). Based on John Stuart Mill's rules of inductive reasoning from 1856, Evan developed a unified concept of causation which is now the generally accepted means for identifying cause-effect relationships in modern epidemiology. Evan's unified concept of causation includes the following criteria:

1. The proportion of individuals with disease should be higher in those exposed to the putative cause than in those not exposed.
2. Exposure to the putative cause should be more common in cases than in those without the disease.
3. The number of new cases should be higher in those exposed to the putative cause than in those not exposed, as shown in prospective studies.
4. Temporally, the disease should follow exposure to the putative cause.
5. There should be a measurable biologic spectrum of host responses.
6. The disease should be reproducible experimentally.
7. Preventing or modifying the host response should decrease or eliminate the expression of disease.
8. Elimination of the putative cause should result in lower incidence of disease.

6.3 Hill's criteria

Bradford Hill (1965) elaborated on Evans criteria as part of work that identified smoking as a cause of lung cancer. Hill's criteria are as follows:

1. Strength of association
2. Consistency.
3. Temporality.
4. Dose response relationship.
5. Plausibility and coherence.
6. Experimental evidence.
7. Specificity.
8. Analogy.

Hill's intention was to provide a set of guidelines that could be used to determine if associations are causal, providing the following cautionary statement: 'none of my viewpoints can bring indisputable evidence for or against the cause and effect hypothesis and none can be regarded as *sine qua non* [1].'

[1] *sine qua non*: an essential condition or element

Strength of association

The first criterion is that of strength of association, which is conventionally measured by the risk (or odds) of disease in exposed individuals compared with the risk (or odds) in the unexposed. The rationale here is that strong associations are unlikely to be a result of uncontrolled bias or confounding. Strong associations are usually considered to be risk ratios in excess of 4 or 5. The term 'small' is used in epidemiology to describe risk ratios from observation studies that are less than 2.0, since it is possible that such associations may be due to bias and/or confounding. Obviously, a relative risk of 1.4 is not actually small in magnitude since it indicates a 40% higher rate in the exposed, and would have a significant demographic effect when applied to populations if the exposure was common. Intervention effects of this magnitude would be regarded as highly significant from a public health or clinical perspective if obtained from randomised trials where confounders are dealt with in the randomisation process. As methods and analysis in observational epidemiology continue to improve, smaller relative risks may be accepted as evidence of causation.

Establishment of an overall risk ratio from a number of studies can be achieved by an analytical technique known as meta-analysis, which may vary from the use of mean or median values to fixed- and random-effects models. Combination of data from several studies may produce a statistically significant risk ratio, whereas individual studies may lack sufficient numbers to achieve statistical significance.

Consistency

Consistent findings from several studies that have investigated the strength of association between a risk factor and disease support an argument that the risk factor is causative. Consistency also applies to the existence and pattern of trend in dose response.

Smoking has been associated with lung cancer in at least 29 retrospective and 7 prospective studies. The consistency of this association provides powerful evidence that smoking causes lung cancer.

Temporality

Causes must preceed the effect. For example, if severe angina due to coronary heart disease led to reduced physical activity and a sedentary lifestyle in a previously active person, such inactivity (although associated with coronary heart disease in a cross-sectional context) could not be held accountable for it. Longitudinal studies are particularly useful for determining the temporal relationship between possible causative factors and outcomes.

Obesity can be identified as a strong risk factor for the incidence of adult onset diabetes in longitudinal studies. However, since the management of adult onset usually results in weight loss, cross-sectional studies of obesity and diabetes may reveal no association, or even a negative one.

Dose response

A dose response effect implies that the likelihood or severity of the outcome is greater with a higher close of the exposure. This may manifest as a comparison between outcomes at multiple levels of exposure. Trends may be linear or curvilinear. While differences between individual

exposure groups may not be statistically significant, the trend across three or more groups may be significant. When several studies are considered it is usual to report what proportion demonstrates a significant trend, and whether significant trends are in the same direction.

Plausibility and coherence

Here we ask if a causal interpretation fits with known facts of natural history and biology of disease, that is does a causal relationship make 'biological sense'? Biological plausibility provides a strong argument for causation, if it is present. However, its absence need not exclude causality, particularly if little is known about the disease under investigation.

Experimental evidence

As outlined above, it is not generally possible to perform experiments on humans in which possible disease-producing agents are administered or risk behaviors encouraged. However, evidence from controlled randomised trials of interventions provide a good argument for causation.

The beneficial effects of a high fibre diet in the prevention of colon cancer in high-risk populations suggests that low fibre intake is a causative factor in disease occurrence.

Reversal of signs of scurvy with vitamin C suggests that vitamin C deficiency is the cause of scurvy.

Specificity

This criteria states that a single exposure generally causes a single disease. This is a hold-over from the concepts of causation that were developed for infectious diseases, though there are many exceptions (e.g. smoking is associated with lung cancer as well as many other diseases). When present, specificity provides evidence of causality, but its absence does not preclude causation.

Analogy

This criteria asks if a similar relationship been observed with another exposure and/or disease (e.g. BSE and scrapie/transmissible mink encephalopathy). This is one of the weakest criteria for causation but it is useful in speculation of how putative causative factors may operate in different contexts.

6.4 Causal web models

Causal factors act in a hierarchical fashion and for this reason it is useful to develop path models to describe and explain the relationships between sufficient and necessary causes of disease. Figure 26 is a path web model showing factors associated with pneumonia in lambs.

Path models are useful because they provide a representation of 'the big picture' — a framework for thinking about the relationships between causes (particularly temporal relationships) and how they interelate with each other. This provides a useful starting point for designing research strategies to investigate diseases that are of interest.

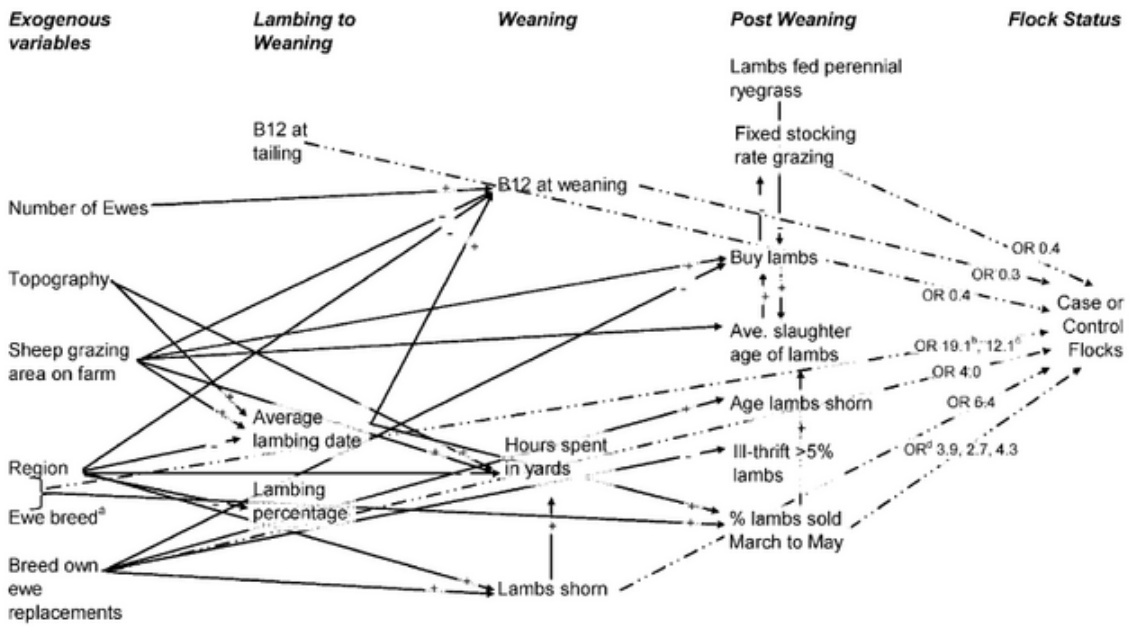


Figure 26: Path model of factors associated with pneumonia in New Zealand lambs. Reproduced from Goodwin-Ray et al. (2008).

7 Sampling

By the end of this unit you should be able to:

- Explain the key features of simple random sampling, systematic random sampling, stratified random sampling, and cluster sampling. Describe the advantages of disadvantages of each approach.
- Describe ways to reduce error when making inferences from sampled data.
- Given the appropriate formulae, calculate the required sample size when you want to estimate a population total, mean, or proportion.
- Given the appropriate formula, calculate the sample size required to detect the presence of disease in a population. Adjust this estimate to account for a test that is imperfect.

Epidemiologists frequently examine populations to:

- Detect the presence of a disease;
- Demonstrate that a disease is not present within a population; and
- Establish the level of occurrence of a disease within a population.

To produce accurate estimates of disease we must be able to measure populations effectively. The exact level of disease within a population will be obtained if every individual within the population is examined (and if there was no measurement error). This technique is a census. However, in many situations a census is impossible and/or excessively expensive. Usually an accurate estimate can be obtained by examining some of the animals (a sample) from the population.

7.1 Probability sampling methods

A probability sample is one in which every element in the population has a known non-zero chance of being included in the sample.

Simple random sampling

Simple random sampling occurs when each subject in the population has an equal chance of being selected.

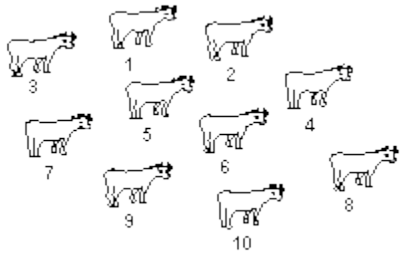


Figure 27: Simple random sampling. If a sample of five cows was required, five random numbers between 1 and 10 would be generated and cows selected on the basis of the generated random numbers.

Systematic random sampling

With systematic random sampling, the selection of sampling units occurs at a predefined equal interval (known as the sampling interval). This process is used when the total number of sampling units is unknown at the time of sampling (e.g. in a study where patients that enter an emergency department of a hospital on a given day are to be sampled — at the start of the study day we do not know the total number of patients seen by the end of the day).

Suppose we are studying inpatient medical records on an ongoing basis for a detailed audit. The total number of records in the population is not likely to be known in advance of the sampling since the records are to be sampled on an ongoing basis (and so it would not be possible to use simple random sampling). However, it would be possible to guess the approximate number of records that would be available per time period and to select a sample of one in every k records as they become available.

Say we require a total of 300 records over a 12-month period to complete the study. If there are, on average, ten new discharge records available per day then total number of records available per year is estimated to be $10 \times 365 = 3650$. To obtain the required number of records per year in the sample, the sampling interval k will be $3650 \div 300 = 12$. Thus, we would take a sample of 1 from every 12 records.

One way to implement this procedure is to identify each record as it is created with a consecutive number. At the beginning of the study a random number between 1 and 12 is chosen as the starting point. Then, that record and every twelfth record beyond it is sampled. If the random number chosen is 4, then the records in the sample would be 4, 16, 28, 40, 52, and so on.

Stratified random sampling

Stratified sampling occurs when the sampling frame is divided into groups (strata) and a random sample taken from within each stratum. Stratified sampling is frequently undertaken to ensure that there is adequate representation of all groups in the population in the final sample. The simplest form is proportional stratified random sampling, where the number sampled within each stratum is proportional to the total number within the stratum.

Suppose that you wish to determine the prevalence of disease in the pig population of a region. Previous surveys have indicated that 70% of the regions pigs are located in very large, intensive specialised pig farms, 20% of pigs are found within smaller farming units (frequently as a secondary enterprise on large dairy farms), and 10% of pigs are kept singly within small plots around towns (by people whose major occupation is not farming). With proportional stratification, a sample would be selected at random from within each stratum such that the aggregated sample would consist of 70% pigs obtained from the large intensive farms, 20% pigs obtained from the smaller pig farms, and 10% pigs obtained from small plots near towns.

If the population can be divided into logical strata whereby the variation within each stratum is small compared with the variation between strata, stratified random sampling will provide a more precise estimate of the population parameter.

We wish to determine average total lactation milk volume (total litres) produced by dairy cows in a region. The region contains two breeds of cattle. One breed (Friesian) is characterised by production of large volumes of milk with low concentrations of milk solids. The other breed (Jersey) is characterised by production of small volumes of milk with high concentrations of milk solids. By dividing the population into breed strata and sampling within each stratum, the average lactation milk volume production of each breed can be estimated with accuracy. The mean milk production for cows within the region can be estimated by working out the weighted mean based upon each stratum mean and stratum size.

Cluster sampling

Cluster sampling occurs when the sampling frame is divided into logical aggregations (clusters) and a random selection of clusters is performed. The individual sampling units (known as

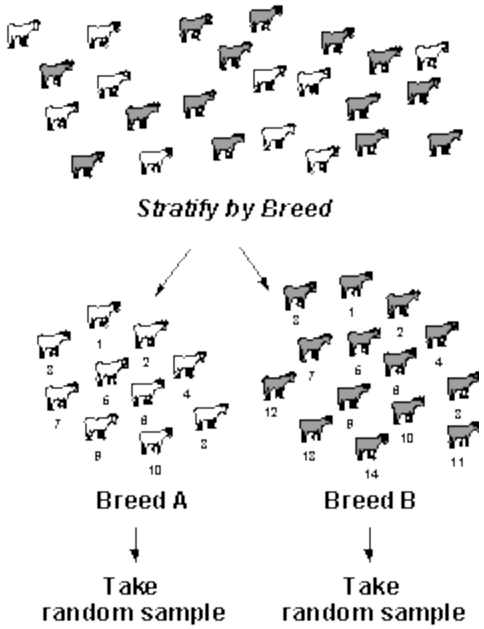


Figure 28: Stratified random sampling. A group of animals are stratified by breed and a random sample within each breed taken.

primary sampling units) within the selected clusters are then examined. Clustering may occur in space or time. For example, a litter of piglets is a cluster formed within a sow, a herd of dairy cows is a cluster within a farm, and a fleet of fishing boats is a cluster formed within space (that is, a port or harbour).

The standard errors of population estimates derived from cluster sampling are often high compared with those obtained from simple random or stratified random sampling procedures. The reason for this is that units within the same cluster tend to be more homogenous than those units from different clusters.

There are two types of cluster sampling:

- One stage cluster sampling occurs when clusters are selected by simple random sampling and then, once selected, all of the listing units within the cluster are examined.
- Two stage cluster sampling occurs when clusters are selected by simple random sampling and then, once selected, a random sample of listing units within each cluster are selected for examination. Estimation of population characteristics is straightforward in this situation when each cluster has the same number of listing units. Estimation of population characteristics is not straightforward when each cluster contains different numbers of listing units (in this case, you will need to consult a statistician).

The number of clusters to sample and the number of listing units within each cluster to sample will depend upon the relative variation of the factor of interest between clusters, compared with within clusters, and the relative cost of sampling clusters compared with the cost of sampling individual listing units.

- When the between-cluster variation is large relative to the within-cluster variation, you will have to sample many more clusters to get a precise population estimate.
- When the between-cluster variation is small relative to the within-cluster variation, you will have to sample many more individual listing units within each cluster to get a precise population estimate.

7.2 Non-probability sampling methods

Non-probability sampling occurs when the probability of selection of an individual within a population is not known and some groups within the population are more or less likely than other groups to be selected. Non-probability sampling methods include:

- Convenience sampling: where the most accessible or amenable sampling units are selected;
- Purposive sampling: where the most desired sampling units are selected; and
- Haphazard sampling: where sampling units are selected using no particular scheme or method. Inherent in this type of sampling is the problem that subconscious forces may influence the person selecting the units in an attempt to 'balance' the sample. For example, a young animal may be preferred for the next selection immediately after an older animal has been selected.

Non-probability sampling will produce biased population estimates, and the extent of that bias cannot be quantified.

7.3 Sampling techniques

Random sampling means that each unit of interest within the population has the same probability of selection into the sample as every other unit. The probability of selection of individual units must not differ. This is irrespective of accessibility, ease of collection or other differences that may exist between individuals. There are several important considerations to take into account before collecting a random sample:

- The target population must be defined and identified.
- A study group (sometimes called a study population) that is representative of the target population must be identified. The study group should not differ in composition from the target population.
- A sampling frame is produced. The sampling frame provides a means for identifying every unit of interest (sampling unit) within the study group.
- Sampling units are selected from the sampling frame using a random (probabilistic) approach such that each sampling unit within the sampling frame has an equal chance of being selected.

Methods of randomisation

There are two principal techniques for random sampling, physical randomisation and the use of random numbers. Physical randomisation is a process where sampling units are selected using physical systems that contain random elements. These include the selection of numbered marbles from a bag, the use of a die, or the toss of a coin.

Random numbers are a sequence of numbers comprising individual digits with an equal chance that any number from 0 to 9 will be present. Tables of random numbers can be used for sample selection. Some computer programs can generate random numbers. These programs use algorithms to produce the sequence of numbers. The sequence of numbers that is generated depends upon the value chosen as the starting value for the algorithm (the seed value). Whilst there is an equal probability that any digit from 0 to 9 will be present in a position chosen at random from the sequence, the actual digit present at each point of the sequence is determined by the seed value. In other words, the exact sequence of random numbers can be reproduced if the process is repeated using the same seed value. Computer-generated random numbers are frequently called pseudo-random numbers for this reason.

Replacement

Samples may be taken in one of two ways: with replacement or without replacement. In sampling with replacement, each selected unit is examined and recorded and then returned to the sampling frame. These units may then be selected into the sample again.

In sampling without replacement, each selected unit is examined and recorded and then withdrawn from the sampling frame. These units are excluded from selection into the sample again. Intuitively, sampling without replacement is the most logical — it is better to have different information from new animals as opposed to having copies of information obtained from the repeated sampling of a single animal. However, there are statistical reasons why sampling with replacement may be employed in certain circumstances. These reasons relate to the mathematics of the estimation process. In sampling with replacement the probability of selection of a unit remains the same from the first selection through to the last selection. The distribution of results within the final sample is described by the binomial distribution. In sampling without replacement, the probability of selection of the next unit changes each time a selection is made. This is due to a reduction in size of the denominator as each unit is drawn. The distribution of results is described by the (more complex) hypergeometric distribution.

The difference between the two sampling procedures is not important when samples are drawn from large populations. Often, the binomial distribution is used to approximate the hypergeometric distribution when analysing the results of samples drawn without replacement from large populations.

Probability proportional to size sampling

When clusters differ widely with respect to the number of units that they contain, unequal probability sampling of clusters will often result in estimates of population characteristics, especially population totals, that have lower standard errors than those obtained from sampling clusters with equal probability. Probability proportional to size sampling avoids these problems.

Say, for example, suppose you need to take a random sample of three herds from a list of the 10 herds shown Table 13. First divide the total population (6700) by the number of herds to be selected (3) to obtain a sampling interval ($6700 \div 3 = 2233$). Next choose a random number between 1 and 2233. Suppose the chosen number is 1814. This should be fitted in position in the list to identify the first herd in the sample. Since 1814 lies between 1601 and 1900, the first selected cluster is herd 4. Next, add the sampling interval to the initial random number: $1814 + 2233 = 4047$. The next cluster to be selected is herd number 6. Add the sampling interval again: $4047 + 2233 = 6280$ and herd 10 is chosen.

Table 13: A cumulative list of herd sizes.

Herd	n	Cumulative n
1	1000	1000
2	400	1400
3	200	1600
4	300	1900
5	1200	3100
6	1000	41000
7	1600	5700
8	200	5900
9	350	6250
10	450	6700

Note that when this technique is used it is possible for the same cluster to be selected twice if the cluster has a population size that is greater than the sampling interval. This is unlikely to happen if the proportion of clusters selected is small, unless one cluster is very much larger than the others. If this occurs, you should select two subsamples of subjects from within the cluster. It is not valid to select another cluster instead, or to repeat the sampling procedure until no clusters are repeated, since either of these two approaches invalidates the required probabilities.

If no estimate of cluster sizes is available it will be impossible to carry out selection proportional to size and clusters must be selected by simple random sampling methods. If this is the case, responses will need to be weighted in any analyses that are undertaken. This requires a count of the total number of sampling units in each selected cluster.

7.4 Sample size

The choice of sample size involves both statistical and non-statistical considerations. Non-statistical considerations include the availability of time, money, and resources. Statistical considerations include the required precision of the estimate, and the variance expected in the data. In descriptive studies we need to specify the desired level of confidence that the estimate obtained from sampling is close to the true population value $(1 - \alpha)$. In analytical studies we may also be interested in the power $(1 - \beta)$ of the study to detect real effects.

Simple and systematic random sampling

The following formulae may be used to derive sample sizes appropriate to estimate population parameters (population total, mean, and proportion) on the basis of a simple random sample.

$$\text{Total:} n \geqslant \frac{z^2 SD^2}{\epsilon^2} \tag{17}$$

$$\text{Mean:} n \geqslant \frac{z^2 SD^2}{\epsilon^2} \tag{18}$$

$$\text{Proportion:} n \geqslant \frac{z^2 (1 - P_y) P_y}{\epsilon^2} \tag{19}$$

Where:

z: the reliability coefficient (e.g. $z = 1.96$ for an alpha level of 0.05)
SD: the population standard deviation of the variable of interest
ϵ: the maximum absolute difference between the sample estimate and the unknown population value
P_y: the unknown population proportion

We want to estimate the mean bodyweight of deer on a farm. We anticipate the standard deviation of body weight in farmed deer of this age is around 30 kg. We would like to be 95% certain that our estimate is within 10 kg of the true mean. How many deer should we include in our sample?

$z = 1.96$
$SD = 30$
$\epsilon = 10$

$n = (z^2 \times SD^2) \; / \; \epsilon^2$
$n = (1.96 \times 1.96 \times 30 \times 30) \div (10 \times 10)$
$n = 34$

We need to sample 34 deer.

We want to estimate the sero-prevalence of brucellosis in a population of cattle. The expected prevalence is 15% and we would like to take enough samples to be 95% sure that our estimate is within 20% of the actual prevalence of disease. How many cattle should be included in our sample?

$z = 1.96$
$P_y = 0.15$
Absolute error $= 0.20 \times 0.15 = 0.03$

$n = [\, z^2 \times (1 - P_y) \times P_y \,] \; / \; \epsilon^2$
$n = [\, 1.96 \times 1.96 \times (1 - 0.15) \times 0.15 \,] \div (0.03 \times 0.03)$
$n = 544$

We need to sample 544 cattle.

Sampling to detect disease

Veterinarians are frequently asked to test groups of animals to confirm the absence of disease. The number of animals that should be tested to provide a specified level of confidence that disease is detected is given by:

$$n = (1 - \alpha^{\frac{1}{D}}) \times (N - \frac{D - 1}{2}) \tag{20}$$

Where:

N: the population size
α: 1 - confidence level (usually $\alpha = 0.05$)
D: the estimated minimum number of diseased animals in the group (that is, population size $\times$ the minimum expected prevalence)

What is the approximate number of animals that should be tested in a herd of 200 to be 95% confident that at least one disease animal will be found if the expected prevalence is 20%?

N = 200
$\alpha = 0.05$
D = 0.20 $\times$ 200 = 40

n = (1 - $\alpha^{1/D}$) $\times$ (N - [D - 1] / 2)
n = (1 - $0.05^{1/40}$) $\times$ (200 - [40 - 1] / 2)
n = 0.072 $\times$ 180.5
n = 13

A minimum of 13 animals need to be tested.

The above formula assumes that the test being used has perfect ability to detect an animal as diseased, if it really is (that is, the test has a sensitivity of 1.0). The number to be tested can be adjusted to account for an imperfect testing procedure, by multiplying the result from the 'standard' equation by the reciprocal of the test sensitivity.

See Chapter 8 for details.

In the example above, we worked out that 13 animals need to be tested to be 95% confident that at least one disease animal will be found if the expected prevalence of disease was 20%. How many animals should be tested if sensitivity of the diagnostic test is 0.90?

n = 13
Se = 0.90

n' = n $\times$ (1 / Se)
n' = 13 $\times$ (1 / 0.90)
n' = 15

Using a diagnostic test with a sensitivity of 0.90, a minimum of 15 animals need to be tested.

8 Diagnostic tests

By the end of this unit you should be able to:

- Explain what is meant by the terms sensitivity, specificity, and positive and negative predictive value as applied to diagnostic tests.
- Given the appropriate formula and test results presented in a 2×2 table, calculate and interpret test sensitivity, specificity, positive and negative predictive value.
- Explain the difference between apparent prevalence and true prevalence. Given data presented in a 2×2 table, calculate apparent and true prevalence.
- Explain what is meant by parallel and series interpretation of diagnostic tests. Provide examples of where parallel and series test interpretation is used (or would be useful).
- Explain how you would estimate the pre-test probability of the various diseases you might encounter in clinical practice.
- Using a nomogram and an estimate of the pre-test probability of disease, determine the the post-test probability of disease in an individual.

A test may be defined as any process or device designed to detect (or quantify) a sign, substance, tissue change, or body response in an animal. Tests included:

- Routine examination of an animal or premises.
- Questions posed during history taking.
- Clinical signs.
- Laboratory findings — haematology, serology, biochemistry, histopathology.
- Post mortem findings.

If tests are to be used in a decision-making context, the selection of an appropriate test should be based on its ability to alter your assessment of the probability that a disease does or does not exist.

8.1 Screening versus diagnosis

In clinical practice, tests tend to be used in two ways. **Screening** tests are those applied to apparently healthy members of a population to detect the presence of disease, disease-causing agents, or subclinical disease. Usually, those animals that return a positive to such tests are subject to further in-depth diagnostic work-up. **Diagnostic** tests are used to confirm or classify disease status, provide a guide to selection of treatment, or provide a prognosis. In this setting, all animals are 'abnormal' and the challenge is to make a correct diagnosis.

8.2 Sensitivity and specificity

The analytic sensitivity of an assay for detecting a given chemical compound refers to the lowest concentration the test can detect. Analytic specificity refers to the capacity of the test to react to a single chemical compound.

Epidemiologic sensitivity and specificity depend on analytic sensitivity and specificity, but are entirely different concepts. Epidemiologic sensitivity answers the question: 'Of all individuals that actually had disease X, what proportion tested positive? Epidemiologic specificity answers the question: 'Of all individuals that were free of disease X, what proportion tested negative? Figure 29 presents this concept diagramatically.

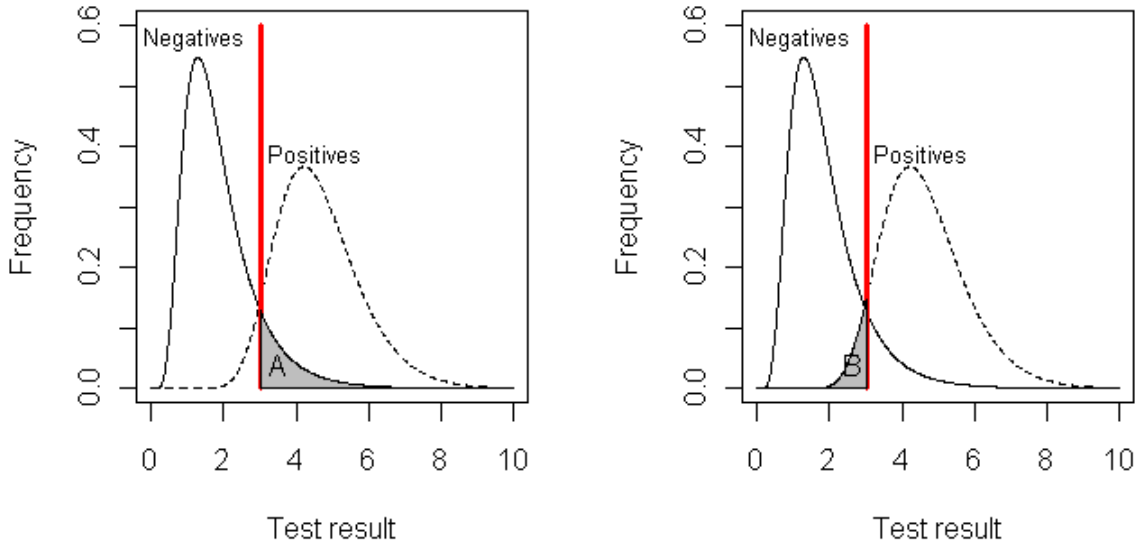


Figure 29: Test results measured on a continuous scale, showing the distribution of results that might be obtained for healthy and diseased individuals. The cut-off value for the test is shown by the vertical solid line: those individuals with a result less than the cut-off value are diagnosed as non-diseased, those individuals with a result greater than the cut-off value are diagnosed as diseased. Using this diagnostic test, disease-negative individuals with a test result greater than the cut-off value ('A' in the left-hand plot) will be false positive. Disease-positive individuals with a test result less than the cut-off value ('B' in the right-hand plot) will be false negatives.

8.3 Accuracy and precision

The accuracy of a test relates to its ability to give a true measure of the substance being measured. To be accurate, a test need not always be close to the true value, but if repeat tests are run, the average of the results should be close to the true value. An accurate test will not over- or under-estimate the true value. Results from tests can be 'corrected' if the degree of inaccuracy can be measured and the test results adjusted accordingly.

The precision of a test relates to how consistent the results of the test are. If a test always gives the same value for a sample (regardless of whether or not it is the correct value), it is said to be precise.

Accuracy

Assessment of test accuracy involves running the test on samples with a known quantity of substance present. These can be field samples for which the quantity of substance present

has been determined by another, accepted reference procedure. Alternatively, the accuracy of a test can be determined by testing samples to which a known quantity of a substance has been added. The presence of background levels of substance in the original sample and the representativeness of these 'spiked' samples make this approach less desirable for evaluating tests designed for routine field use.

Precision

Variability among test results might be due to variability among results obtained from running the same sample within the same laboratory (repeatability) or variability between laboratories (reproducibility). Regardless of what is being measured, evaluation of test precision involves testing the same sample multiple times within and/or among laboratories.

8.4 Test evaluation

The two key requirements of a diagnostic test are: (1) the test will detect diseased animals correctly, and (2) the test will detect non-diseased animals correctly. To work out how well a diagnostic test performs, we need to compare it with a 'gold standard.' A gold standard is a test or procedure that is absolutely accurate. It diagnoses all diseased animals that are tested and misdiagnoses none.

Histopathological and microbiological examination of the small intestine is generally regarded as the gold standard test for Johne's disease in cattle.

Histopathological examination of the brain stem is the gold standard test for bovine spongiform encephalopathy.

Once samples are tested using the gold standard and the test to be evaluated, a 2 × 2 table can be constructed, allowing test performance to be quantified. The usual format is shown in Table 14.

Table 14: Diagnostic test data presented in a 2 × 2 table format.

	Gold std +	Gold std -	Total
Test positive	a	b	$a+b$
Test negative	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

Sensitivity

The sensitivity of a test is defined as the proportion of subjects with disease that test positive $[p(T^+|D^+)]$. A sensitive test will rarely misclassify animals with the disease. Sensitivity is a measure of accuracy for predicting events.

$$\text{Sensitivity} = \frac{a}{(a+c)} \tag{21}$$

Sensitivity is:

- The conditional probability of a positive test, given the presence of disease.
- The likelihood of a positive test in a diseased animal.
- The proportion of animals with disease that have a positive test for the disease.
- The true positive rate (relative to all animals with disease).

Specificity

The specificity of a test is defined as the proportion of subjects without disease that test negative $[p(T^-|D^-)]$. A highly specific test will rarely misclassify animals that are not diseased.

$$\text{Specificity} = \frac{d}{(b+d)} \tag{22}$$

Specificity is:

- The conditional probability of a negative test, given the absence of disease.
- The likelihood of a negative test in an animal without disease.
- The proportion of animals without the disease that have a negative test for the disease.
- The true negative rate (relative to all animals without disease).

Sensitivity and specificity are inversely related and in the case of test results measured on a continuous scale they can be varied by changing the cut-off value. In doing so, an increase in sensitivity will often result in a decrease in specificity, and *vice versa*. The optimum cut-off level depends on the diagnostic strategy. If the primary objective is to find diseased animals (that is, to minimise the number of false negatives and accept a limited number of false positives) a test with a high sensitivity is required. If the objective is to make sure that every test positive is 'truly' diseased (minimise the number of false positives and accept a limited number of false negatives) the diagnostic test should have a high specificity.

Positive predictive value

The positive predictive value is the proportion of subjects with positive test results which have the disease.

$$\text{Positive predictive value} = \frac{a}{(a+b)} \tag{23}$$

Positive predictive value is:

- The predictive value of a positive test.
- The post-test probability of disease following a positive test.
- The posterior probability of disease following a positive test.

Negative predictive value

The negative predictive value is the proportion of subjects with negative test results which do not have the disease.

$$\text{Negative predictive values} = \frac{d}{(c+d)} \tag{24}$$

Negative predictive value is:

- The predictive value of a negative test.
- The post-test probability of no disease following a negative test.
- The posterior probability of no disease following a negative test.

Predictive values quantify the probability that a test result for a particular animal correctly identifies the condition of interest. Estimation of predictive values requires knowledge of sensitivity, specificity and the prevalence of the disease of interest in the population. The effect of prevalence on predictive values is considerable. Given a prevalence of disease in a population of around 30% and we are using a test with 0.95 sensitivity and 0.90 specificity, the predictive value of a positive test would be 0.80 and the predictive value of a negative test would be 0.98. If the prevalence of disease is only 3% and the test characteristics remain the same, the predictive value of a positive and negative test will be 0.23 and 0.99, respectively.

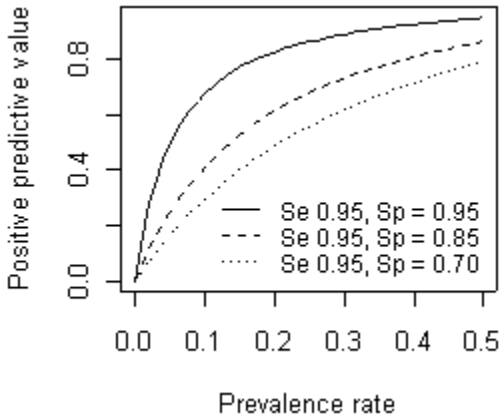


Figure 30: Relationship between prevalence and positive predictive value for tests of different sensitivities and specificities.

Sensitivity and specificity are properties of a test and don't change with prevalence.

If the prevalence increases, positive predictive value increases and negative predictive value decreases. If the prevalence decreases, positive predictive value decreases and negative predictive value increases.

The more sensitive a test, the better its negative predictive value. The more specific a test, the better its positive predictive value.

8.5 Prevalence estimation

The estimate of disease prevalence determined on the basis of an imperfect test is called the **apparent prevalence**. Apparent prevalence is the proportion of all animals that give a positive test result. It can be more than, less than, or equal to the actual proportion of diseased animals, the **true prevalence**. If sensitivity and specificity of a test are known, true prevalence can be calculated using the Rogan and Gladen (1978) formula:

$$p(D^+) = \frac{AP - (1 - Sp)}{1 - [(1 - Sp) + (1 - Se)]} = \frac{AP + Sp - 1}{Se + Sp - 1} \tag{25}$$

Where:

AP: apparent prevalence
Se: sensitivity (0 - 1)
Sp: specificity (0 - 1)

Individual cow somatic cell counts (ICSCC) are used as a screening test for subclinical mastitis in dairy cattle. This test has a sensitivity of 0.90 and a specificity of 0.80. The apparent prevalence of mastitis in this herd using the screening test is 23 cases per 100 cows. True prevalence p(D+) may be calculated as follows:

AP = 0.23
Se = 0.90
Sp = 0.80

p(D+) = (AP + Sp - 1) ÷ (Se + Sp - 1)
p(D+) = (0.23 + 0.80 - 1) ÷ (0.90 + 0.80 - 1)
p(D+) = 0.03 ÷ 0.70
p(D+) = 0.04

The true prevalence of mastitis in this herd is 4 cases per 100 cows.

8.6 Diagnostic strategies

Clinicians commonly perform multiple tests to increase their confidence that a patient has a particular diagnosis. When multiple tests are performed and all are positive, the interpretation is straightforward: the probability of disease being present is relatively high. It is far more likely however, that some of the tests return a positive result and others will be negative. We can deal with this problem by interpreting test results in **parallel** or **series**.

Parallel interpretation

Parallel interpretation means that when multiple tests are run an individual is declared positive if at least one of the multiple tests returns a positive result. Interpreting test results in parallel increases the sensitivity and therefore the negative predictive value for a given disease prevalence. However, specificity and positive predictive value are lowered. As a consequence, if a large number of tests are performed and interpreted in this way then virtually every individual will be considered positive.

Serial interpretation

Series interpretation means that when multiple tests are run an individual is declared positive if all tests return a positive result. Series interpretation maximises specificity and positive predictive value which means that more confidence can be attributed to positive results. It reduces sensitivity and negative predictive value, and therefore it becomes more likely that diseased animals are being missed.

8.7 Screening and confirmatory testing

With a screening and confirmatory test strategy a test is applied to every animal in the population to screen the population for positives. Ideally, this test should be easy to apply and low in cost. It also should be a highly sensitive test so that it misses only a small number of diseased animals. Its specificity should still be reasonable, so that the number of false positives subjected to the confirmatory test remains economically justifiable.

Individuals that return a negative result to the screening test are regarded as true negatives and not submitted to further examination. Animals positive to the screening test are subjected to a confirmatory test. The confirmatory test can require more technical expertise and more sophisticated equipment, and be more expensive, because it is only applied to a reduced number of samples. But it has to be highly specific, so that any positive reaction to the confirmatory test is considered a definitive positive.

During the early phase of disease control programs (e.g. programs to eradicate tuberculosis) the apparent prevalence will be higher than the true prevalence, as a consequence of test specificity being less than 1.00. As the program continues, test positive animals are identified and culled which results in a decrease in true prevalence. As true prevalence declines the positive predictive value of testing declines, increasing the proportion of false positives. At this stage of the control program a highly specific test is required. In some cases it may be necessary to use a number of tests interpreted in series to increase specificity.

SPINS and SNOUTS. SPecific tests are needed to rule a diagnosis IN, and highly SeNsitive tests are needed to rule them OUT. When a disease is rare however (with a prevalence of less than 0.01) the specificity of a test is rarely high enough to give adequate positive predictive value. Only the sensitivity is useful in the rare disease case. To remember this:

> Thinking about,
> SPIN and SNOUT
> In cases where
> Disease is rare.
> Dont use SPIN,
> But keep SNOUT in.

Positive and negative predictive values are more useful to the clinician than sensitivity and specificity. Predictive values vary with prevalence, a common mistake is to assume they are fixed. You need to know the prevalence of disease to derive valid estimates of positive and negative predictive value.

8.8 Likelihood ratios

Diagnostic testing is often undertaken to help us decide whether or not an individual is diseased. Because diagnostic tests are imperfect (that is, false positives and false negatives occur) we

should move away from the 'test positive = disease positive', 'test negative = disease negative' paradigm and think about testing as a process that provides us with a probability estimate of the presence of disease in an individual. Likelihood ratios provide a means for doing this.

The likelihood ratio for a positive test tells us how likely we are to find a positive test result in a diseased individual compared with a non-diseased individual. The likelihood ratio for a positive test is estimated on the basis of dividing the probability of a particular test result in the presence of disease (sensitivity) by the probability of the test result in the absence of disease (1 - specificity). The likelihood ratio for a negative test equals (1 - sensitivity) divided by the specificity. Thus:

$$LR^{+} = \frac{Se}{1 - Sp} \tag{26}$$

$$LR^{-} = \frac{1 - Se}{Sp} \tag{27}$$

Where:

Se: sensitivity (0 - 1)
Sp: specificity (0 - 1)

Likelihood ratios (LR) can be calculated using single cut-off values, so we have one pair of likelihood ratios: one for a positive (LR+) and another for a negative test result (LR-). More information can be extracted from the diagnostic test by using multilevel likelihood ratios. In this case ranges of test results will have associated likelihood ratio values.

Likelihood ratios provide a quantitative measure of the diagnostic information contained in a particular test result. If we know the prior likelihood that an animal has a certain condition (= pre-test odds of disease) the likelihood ratio of the test multiplied by the prior likelihood gives us an estimate of the post-test likelihood of disease (= post-test odds). This result can be re-expressed as a probability to make it more interpretable. The relationship between odds and probability is as follows:

$$\text{Odds of event} = \frac{\text{Probability of event}}{\text{1 - Probability of event}} \tag{28}$$

$$\text{Probability of event} = \frac{\text{Odds of event}}{\text{1 + Odds of event}} \tag{29}$$

Individual cow somatic cell counts (ICSCC) are used as a screening test for sub-clinical mastitis in dairy herds. A client has a herd of dairy cows where the (true) prevalence of subclinical mastitis is estimated to be around 5%. Your herd testing authority tells you that the positive likelihood ratio for a cell count of 300 – 400 cells/mL is 14.50.

You are called to examine an individual cow from this herd and find that she has an ICSCC of 320,000 cells/mL. What is the probability that this cow really has mastitis? The posterior probability of mastitis in this cow is determined as follows:

1. The pre-test probability of mastitis: 50 ÷ 1000 = 0.05.
2. The pre-test odds of mastitis: 0.05 ÷ (1 - 0.05) = 0.053.
3. The post-test odds of mastitis given a positive test result: pre-test odds × LR(+) = 0.053 × 14.5 = 0.76.
4. The post-test probability of mastitis given a positive test result: 0.76 ÷ (1 + 0.76) = 0.43.

The post-test probability of a cow with a ICSCC of 320,000 cells/mL being mastitic is around 43%.

Post-test probabilities can be quickly determined in practice by using a nomogram (Figure 31). On the left hand side of the nomogram we mark the pre-test probability that the individual being examined has disease. We next identify the point defining the likelihood ratio of a positive test result along the middle scale. Finally, we draw a straight line from the pre-test probability estimate through the likelihood ratio value to the corresponding post-test probability value on the right-hand side of the chart.

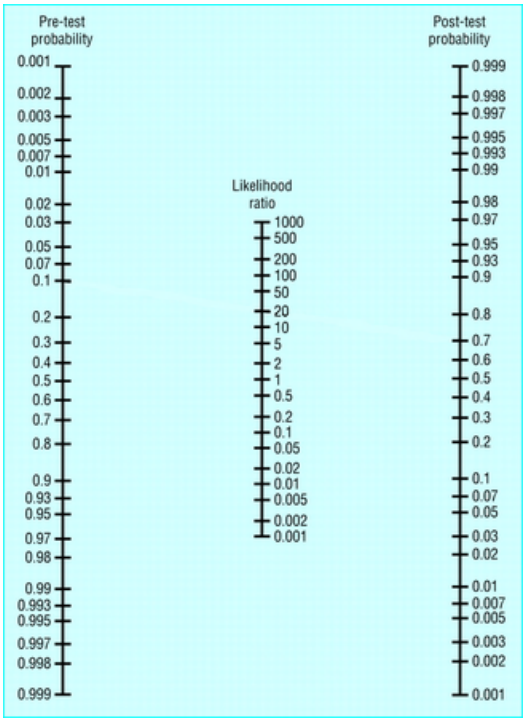


Figure 31: Nomogram for post-test probability calculations using likelihood ratios of a positive test result.

A nice feature of this approach to evaluating test information is that sequential testing can be easily handled. If we are using serial interpretation, the post-test probability of disease from the first test becomes the pre-test probability for the second test.

To continue the mastitis example described above lets imagine that we examine our cow and as part of that examination we test milk from each quarter using a rapid mastitis test (RMT). We are told that the sensitivity and specificity of the RMT is 0.70 and 0.80, respectively. Our cow returns a positive result to the RMT. What is the probability of this cow being mastitic, given this additional information?

The likelihood ratio of a positive RMT is (0.70 / 1 - 0.80) = 3.5. If the pre-test probability of disease is 43% we can use a nomogram to estimate the posterior probability of disease, given a positive test, as 72%. We are now 72% certain that this cow has mastitis.

The advantage of the likelihood ratio method of test interpretation is that we can better appreciate the value (i.e. the increase in post-test probability) provided by each diagnostic test that is applied (in the above example, ICSCC provided more information compared with the RMT). If the cost of each test applied is known the cost per unit increase in post-test probability can be determined, enabling us to be more objective in our use of diagnostic resources.

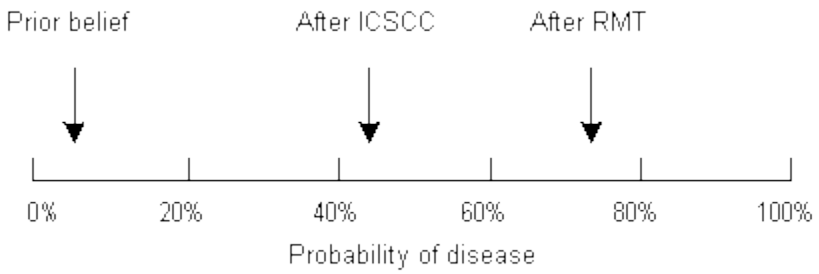


Figure 32: Diagram showing how the estimated probability of disease changes after applying a series of diagnostic tests. In our example of the cow with mastitis, we had a prior belief that the probability of the cow being mastitic was 5%. After considering the ICSCC result this probability increased to 43%. After applying a rapid mastitis test and getting a positive result, the probability of the cow having mastitis increased to 72%.

9 Outbreak investigation

By the end of this unit you should be able to:

- Describe the steps to take during an outbreak investigation, including description of the outbreak by animal, place and time.
- Explain why it is important to establish a case definition when investigating a disease outbreak.
- List methods you might use to enhance surveillance once an outbreak of disease has been identified

An outbreak is a series of disease events clustered in time. During an outbreak the investigator asks the questions:

- What is the problem?
- Can something be done to control it?
- Can future occurrences be prevented?

These notes outline an approach to investigating outbreaks of disease in animal populations. Although the term outbreak implies a sudden (and possibly spectacular) event (e.g. an outbreak of botulism in feedlot cattle), be aware that outbreaks can be of a more insidious nature: some causing subclinical losses in a population of animals over an extended period before being identified, characterised and investigated.

9.1 Verify the outbreak

What is the illness?

Once a suspected outbreak is identified, identifying the specific nature of the illness is an important early step. An attempt should be made to characterise cases (leading towards a formal case definition, see below). It may not be possible to make a definitive diagnosis at this stage. What is required is a working definition of the disease or syndrome: for example 'ill thrift in recently weaned calves' or 'sudden death in grower pigs.'

Is there a true excess of disease?

The first issue to be certain of is whether or not the outbreak is genuinely an unusual event worthy of special attention. The number of cases per unit time should be substantially greater than what is normal for the group of individuals under investigation. It is common to have owners and others concerned about a possible outbreak which is transient increase in the normal level of endemic disease.

9.2 Investigating an outbreak

Establish a case definition

A case definition is the operational definition of a disease for study purposes. A good case definition has two parts: (1) it specifies the characteristics of the population at risk, and (2)

it specifies what distinguishes cases from other members of the population. A case definition ensures that the outcome of interest is consistently defined across space (e.g. among different investigation centres in a large scale outbreak) and over time. Consistency in case definition is important since it will allow the incidence of disease to be measured which in turn allows responses to control efforts to be monitored.

In an outbreak of this severe and often fatal pneumonia in delegates attending the 58th annual meeting of the American Legion, Department of Pennsylvania a case was considered Legionnaires' disease if it met clinical and epidemiologic criteria. The clinical criteria required that a person have onset between 1 July and 18 August 1976, an illness characterised by cough and fever (temperature of 38.9° or higher) or any fever and chest x-ray evidence of pneumonia. To meet the epidemiologic criteria, a patient either had to have attended the American Legion Convention held 21 – 24 July 1976, in Philadelphia, or had to have entered Hotel A between 1 July 1976 and the onset of illness.

Reference: Fraser DW, Tsai TR, Orenstein W, Parkin WE, Beecham HJ, Sharrar RG, Harris J, Mallison GF, Martin SM, McDade JE, Shepard CC, Brachman PS (1977). Legionnaires' disease — description of an epidemic of pneumonia. New England Journal of Medicine, 297:1189-1197.

Enhance surveillance

When it is suspected that an outbreak is occurring, enhanced surveillance can be useful to identify additional cases. Enhanced surveillance may involve both heightening awareness to increase passive case reports and implementing targeted surveillance. Techniques include directly contacting field practitioners by telephone, facsimile or email, via health department web pages and email discussion groups. For large outbreaks media releases (print, television, radio) can be extremely effective.

Describe outbreak according to individual, place and time

Record the physical layout of the farm premises. Draw and label a map of all pens, pastures or other physical characteristics that demarcate groups of animals. Identify the pens or areas using the producer's system (otherwise the results from groups may be meaningless to all parties). If feeds are involved, indicate storage locations. Record the current numbers of animals in each pen or paddock and the intended maximum capacity of these areas. These numbers are useful for defining denominators when you want to calculate incidence or prevalence.

Record the animal 'calendar' intended by management for the relevant animals. The animal calendar is what happens to animals as they move through their production cycle. These are the what's and when's (what age or days in production cycle) for being moved, fed, vaccinated or otherwise handled (e.g. dried off, dehorned, bred, examined for pregnancy) and otherwise exposed to potential risk factors. The easiest way to do this is to start at a point in the production cycle and work your way around the cycle, recording the management policy for each event. Double-check later with the people who actually carry out these procedure to see if this is indeed what happens to them.

Record the policy that determines when or why animals are routinely moved from one group to another during normal operation. For example, the policy for drying cows off may be that lactating cows are dried off when they fall below a particular production level or when they reach so many days before calving (whichever comes first). A related issue is how often and on what schedule is production assessed and cows moved. For example, is this done on a certain day of the week, monthly after testing or when? If investigating a calf scour problem, the policy

for moving cows from the high string to the low string is not relevant but the policy for moving cows during the dry period may be.

For those policies that may be related to the problem at hand compare evidence of the actual practice to the management policy. Has this policy been enforced well, is enforcement variable, or is the policy becoming lax? Continuing the example above, determine if the dry off policy is actually practiced by assessing the lengths of dry periods of recently calved cows. Management often has one policy (e.g. calves are weaned at 30 days) but employees may be executing another.

Obtain specific information on practices. For example, if the transmission of an infectious agent by treatment equipment is potentially involved, ask how and when the equipment is sterilized. Between animals? Is it washed with soap first? What concentration of what disinfectant is used? How long is the contact time?

Collect historical, clinical and productivity data on those individuals that are affected (cases) and those that are not affected (non-cases). If possible, all cases of diseased animals should be included in the investigation. If there are large numbers of unaffected individuals you may select a representative sample of unaffected individuals for examination (controls). You may consider matching controls with some characteristic of the cases e.g. age and gender.

Plot an epidemic curve by identifying the first detected case (index case) and then graphing subsequent numbers of cases per day or per week from the index case through to the end of the outbreak. Does the epidemic show common source or propagated properties?

Develop hypotheses about the nature of exposure

At this stage, you will probably have some suspicions about what has caused the outbreak — that is, you will have started to form some hypotheses. Your next job is to test these hypotheses using the various analytical techniques described below.

Conduct analytical studies

Part of the data collection procedure above will have entailed collecting individual-level details such as age, sex, breed, date of parturition, stage of production. Individuals should be categorised according to the presence of each attribute. Attack rate tables divide the cohort of interest into exposed and non-exposed groups. Attack rates are then calculated for each exposure by dividing the number diseased by the group size (Table 15).

Table 15: Attack rate table for an outbreak of food poisoning.

Food	Exposed				Unexposed			
	Ill	Well	Total	AR (%)	Ill	Well	Total	AR (%)
Ham	36	5	41	88	2	11	13	15
Salad	40	4	44	91	9	6	15	60
Prawns	16	15	31	52	10	13	23	43

The exposure which is most likely to have served as a vehicle for an outbreak is that with the greatest difference in attack rate for exposed and unexposed individuals. In Table 15 the greatest difference in attack rate is for the ham-exposed and ham-unexposed groups. This would support the hypothesis that ham was the source of this outbreak of food poisoning. An alternative is to calculate the risk ratio of disease for each exposure. Essentially this is the attack rate for the exposed individuals divided by the attack rate for unexposed individuals — the exposure with the highest risk ratio being the likely vehicle for the outbreak. It is also useful to calculate the population attributable fraction for each exposure. This will identify the percent of the risk of disease in the exposed group that is due to exposure. The closer this value is to 100% the more likely the exposure accounted for the outbreak.

Dozens become sick from food tainted by salmonella

Three sent to hospital with severe symptoms

SCOTT ROBERTS
STAFF REPORTER

At least 75 people are sick and three are in hospital after eating a Mother's Day buffet tainted with salmonella bacteria last Sunday at the Royal Botanical Gardens in Burlington.

About 300 people attended the brunch. Health officials have only been able to contact about 170 and are expecting the number of ailing people to rise.

"This is a very high attack rate for salmonella," said Dr. Bob Nosal, medical officer of health for the Halton Region. "Right now about 40 per cent of those contacted are ill. We usually don't see numbers that high."

Nosal said this strain of salmonella was potent, causing severe diarrhea and vomiting in many victims. Other symptoms include fever, nausea, headaches and abdominal pain.

Public health investigators are trying to pinpoint how the salmonella poisoning occurred, which specific foods were involved and who was to blame. They obtained 14 food samples from the buffet, which have been sent off to the Ontario Public Health Laboratory in Toronto for testing. Results are expected in a few days.

Health officials are also administering detailed questionnaires to those who attended the brunch in hopes of finding the cause. They are also asking for stool samples from those who became ill.

Royal Botanical Gardens contracts out its food services.

Halton Region public health officials would not disclose the name of the catering company responsible for the Mother's Day meal. Royal Botanical Gardens officials were not available for comment last night.

The Halton Region Health Department is not allowing the catering company to prepare any more buffets at the Royal Botanical Gardens until further notice.

"My understanding is that there haven't been significant problems or issues with this catering company in the recent past," Nosal said.

Figure 33: Report of an outbreak of Salmonellosis in humans arising from a contaminated buffet lunch. Source: The Globe and Mail (Toronto, Canada) Thursday 19 May 2005.

9.3 Implement disease control interventions

At this stage it may be possible to produce a hypothesis regarding the cause of the outbreak and to implement controls on the basis of these hypotheses. Provide written and verbal instructions to your client detailing your approach to controlling the outbreak. Gardner (1990b) provides a nice summary of how to write up an outbreak investigation.

Ensure the appropriate measures are being taken to monitor the response to your interventions. This allows you to monitor how things are going and to revise your control plan at the first opportunity if things don't work out as expected. If further investigation is warranted then other epidemiological studies (case-control, prospective cohort etc) may be designed and implemented. You may also use more complex analytical techniques to analyse data that has already been collected (multivariate techniques).

10 Appraising the literature

By the end of this unit you should be able to:

- Describe, in your own words, the four main areas that should be considered when appraising the scientific literature.
- Explain what is meant by the terms internal and external validity.
- Explain the difference between the eligible population and the study population.

Reading the literature is necessary to keep up to date with new developments and to learn more about a particular area of science that interests us.

Fortunately, there appears to be no shortage of literature available to read, and our ability to source this literature easily has been facilitated by the Internet (either in the form of peer-reviewed articles published on-line by established journals or as pre-print publications published by individuals on their own web pages). Although the Internet allows information to be widely disseminated, the quality of that information varies widely. As a result, as good scientists, we need to be discerning about what we read and (more importantly) what we believe. A systematic method of appraising (or evaluating) the literature helps us to do this. These notes outline a systematic approach to appraising the epidemiological literature, which consists of:

1. Describing the evidence;
2. Assessing the internal validity of the study;
3. Assessing the external validity of the study; and
4. Comparing the results with other available evidence.

These notes outline an approach for critically appraising epidemiological studies (i.e. those that investigate the relationship between a set of exposures and a defined outcome). An excellent series of articles providing guidelines for appraising other types of articles appeared in the British Medical Journal in 1997 (Greenhalgh 1997a, 1997b, 1997c, 1997d, 1997e, 1997f, 1997g, and Greenhalgh and Taylor 1997). Much of the technical material from these articles has been compiled into a very readable textbook on the subject by the same author (Greenhalgh 2006).

10.1 Description of the evidence

The first step in evaluating a scientific article is to understand exactly what relationship was being evaluated and what hypothesis was being tested. It should be relatively easy for the reader to identify the exposure variable(s), the outcome variable and the study design (survey, case-control, cohort study, clinical trial). The subjects that were studied in terms of source populations, eligibility criteria, and participation rates of the different groups that are being compared should be clearly stated.

Having defined the topic of study, it is then useful to summarise the main result — what is the result in terms of the association between exposure and outcome? It should be possible to express the main result in a simple table and obtain from the paper the means to calculate the appropriate measure of association (risk ratio, odds ratio, difference in proportions) and the appropriate test of statistical significance.

10.2 Internal validity

Non-causal explanations

Having described the study the next step is to assess its internal validity — that is, are the results valid for the the subjects who were studied? Consider three three non-causal mechanisms that might have influenced the internal validity: **bias**, **confounding** and **chance**. The order of these non-causal explanations is important. If there is severe observation bias, no analytical manipulation of the data will overcome the problem. If there is confounding, then appropriate analysis will (in most cases) overcome the problem. The assessment of chance variation should be made on the main result of the study, after considering issues of bias and confounding.

Causal explanations

Once the non-causal explanations for the results (bias, confounding, and chance) have been ruled out, attention turns to considering the features of the study that support a claim that there is a causal relationship between the exposure and outcome. Five causal mechanisms should be considered:

1. **Is there a correct temporal relationship?** For a relationship to be causal, the putative exposure must act before the outcome occurs. In a prospective study design where exposed and non-exposed subjects are compared, this requirement is established by ensuring that subjects do not already have the outcome of interest when the study starts. The ability to clarify time relationships is weaker in retrospective studies, and care is required to ensure that possible causal factors did in fact occur before the outcome of interest. A difficulty in all study designs, but more so in retrospective studies, is that the occurrence in biological terms of the outcome of interest may precede the recognition and documentation of that outcome by a long and variable period of time (e.g. some cancers).

2. **Is the relationship strong?** A stronger association, that is a larger risk ratio, is more likely to reflect a causal relationship. As a measured factor gets closer to a biological event on the causal pathway, risk ratios become larger. The fact that a relationship is strong does not always mean that the exposure-outcome relationship is causal, however if there is bias it must be large and therefore easy to identify. If a strong relationship is due to confounding, either the association of the exposure with the confounder must be very close, or the association of the confounder with the outcome must be very strong.

3. **Is there a dose-response relationship?** In some circumstances the demonstration of a smooth dose-response relationship may be a strong argument against an identified relationship arising as a result of bias. In general, we should expect uni-directional dose-effect relationships and evidence that this is not the case should be considered carefully.

4. **Consistency?** A causal relationship will be expected to apply across a wide range of study subjects. An association identified in one study that is consistent with the same association identified in a different groups of subjects is supportive of causation. The difficulty with consistency is that very large data sets are required to assess the similarity or otherwise of associations in different subgroups of subjects. Even with adequate numbers, the subgroups to be compared need to be defined on *a priori* grounds.

5. **Specificity?** It has been argued that specificity (that is, a given exposure produces a specified outcome) provides good evidence for causality. Specificity may be useful, if we do not make it an absolute criterion, as one exposure may produce various outcomes, and one outcome may result from various exposures. The concept is often useful in study design: as a check on response bias we may deliberately collect information on factors which we expect to be the same in groups that we are comparing. Similar results across groups will indicate a lack of observation bias.

10.3 External validity

Before discussing external validity it is useful to define the different populations involved in observational epidemiological research:

1. The **study population**: the individuals who actually took part in the study.
2. The **eligible population**: individuals who met the criteria to be included in the study.
3. The **source population**: the population from which eligible study subjects were drawn.
4. The **external population**: individuals that are not part of the source population (e.g. those in another region or country). It is the intention of many studies to generalise the results to an external population.

With external validity we consider how appropriate it is to apply the results to populations apart from the study population.

Can the results be applied to the eligible population?

The relationship between the study population and those who met the study inclusion criteria but did not take part should be well documented. Losses due to non-participation have to be considered carefully as they are likely to be non-random and the reasons for loss may be related to the exposure and/or the outcome.

Can the results be applied to the source and external populations?

The important issue is not whether the individuals who were studied are 'typical' or representative of the source population, but whether the association between outcome and exposure given by the study participants is likely to apply to other groups. In general, the difficulties of applying results from one group of subjects to another will be minimal for issues of basic physiology and maximal for effects in which cultural and psycho-social aspects are dominant.

10.4 Comparison with other evidence

For many clinical questions a large amount of evidence is available which comes from different types of studies. In these circumstances it is useful to consider a hierarchy of evidence. Given that studies are adequately performed within the limitations of the design used, the reliability of the information from them can be ranked (highest to lowest) as follows:

1. Randomised clinical trials.
2. Cohort and case-control studies.
3. Other comparative studies.
4. Case series, descriptive studies, clinical experience.

Randomised clinical trials, if properly performed on adequate numbers of subjects, provide the strongest evidence of causation because of the unique advantages they provide in terms of overcoming problems of bias and confounding.

Consistency

This is the most important characteristic used in the judgement that an association is causal. To say that the study results are consistent requires that the association has been observed in a number of different studies, each of which individually can be interpreted as showing a causal explanation. Variation in study methodology and study populations make it unlikely that the same biases or confounding factors would be present in all of them. Lack of consistency argues against causality.

Plausibility

Plausibility refers to the observed association being biologically understandable on the basis of current knowledge concerning its likely mechanisms. Be aware that any dramatically new observation may be in advance of current biological thinking and its lack of plausibility may reflect deficiencies in biological knowledge rather than error in observation. For example:

- John Snow effectively prevented cholera in London 25 years before the isolation of the cholera bacillus and the general acceptance of the principle that the disease could be spread by water.
- Percival Pott demonstrated the causal relationship between exposure to soot and scrotal cancer some 150 years before the relevant carcinogen was isolated.

Coherency

An association is regarded as coherent if it fits the general features of the distribution of both the exposure and the outcome under assessment. If lung cancer is due to smoking, the frequency of lung cancer in different populations and in different time periods should relate to the frequency of smoking in those populations at earlier relevant time periods.

If the exposure variable under study causes only a small proportion of the total disease, the overwhelming influence of other factors may make the overall pattern inconsistent.

11 Exercise: outbreak investigation

This exercise has been adapted from Gardner (1990b).

A veterinarian in a mixed practice has been investigating an ongoing diarrhoea problem in neonatal pigs in a 150-sow breeding/finishing herd. In the 12 months prior to the outbreak, 7% of litters had diarrhoea but over recent weeks the proportion of litters affected has increased to about 40%. As part of the investigation the veterinarian submitted 3 acutely affected pigs to the regional diagnostic laboratory. Of the 3 pigs, 1 was infected with *E. coli* serotype 08 but other pathogenic bacteria and viruses were not isolated from the other 2 pigs. Lesions in all 3 pigs were consistent with an acute enteritis. The veterinarian asks you to assist.

As background to the problem, the veterinarian provides you with a map showing the layout of the sheds, a description of normal management procedures, and recent records for farrowing sows as detailed below:

11.1 The problem

Shed design. The shed has 16 concrete-floored pens (oriented in a single row in a west - east direction. Pen 1 is near the entrance door at the western end of the shed and pens run in numerical sequence to pen 16 which is located near the extraction fans. The pit underneath the sows is flushed at least twice daily. During the study, pen 14 was under repair and was not used.

Management - treatments. Sows are moved into cleaned and disinfected pens in the farrowing shed on about day 110 of gestation. Sows farrow with minimal supervision. On the first day of life, pigs have their needle teeth clipped and are provided with heat lamps. No vaccines are given to sows or baby pigs for control of enteric disease. Sows are fed *ad libitum* during lactation with a high energy ration (15.5 MJ DE/kg). During gestation, they are fed about 2.0 to 2.5 kg of a lower energy ration plus about 0.5 kg/day of recycled manure for control of enteric infections and parvovirus. Piglets in litters with diarrhoea are treated with oral furazolidone and electrolytes are offered *ad libitum* in shallow bowls in each pen.

Records. Records are provided from a recent set of 26 farrowings (April 2002) for you to examine before your visit. Before April 2002 the records of diarrhoea were insufficiently detailed to be of value in the current investigation.

11.2 Diagnosis

How valid are owner-diagnoses of scours-related deaths? How could you improve their validity in the future?

11.3 Measures of disease frequency

Estimate the following rates from the data:

- The scours-specific mortality rate.
- The proportional mortality rate for scours.

- The case fatality rate for scours.
- The proportion of litters affected with scours.
- The preweaning mortality rate.

11.4 Investigation

Outline your approach to investigating this diarrhoea problem (at this stage there is no need to calculate any factor-specific rates). What initial conclusions or hypotheses did you formulate after examining the history and laboratory findings, and temporal and spatial patterns of disease?

11.5 Measures of association

Analyse the records from the 26 April farrowings and calculate some factor-specific rates or relative risks either by hand or by using computer software available for that purpose. For example:

- What was the risk ratio of scours in parity 1 litters, compared with litters from all other parities?
- What was the risk ratio of scours in litters from sick sows, compared with litters from healthy sows?
- What was risk ratio of scours in large litters, compared with small litters?
- What was the risk ratio of scours in litters born in pens 1 – 8, compared with litters born in pens 9 – 16?

Test the statistical significance of the difference between the two rates in each case. How helpful are the data in allowing you to formulate better hypotheses? Could confounding be a problem and how would you deal with it at this stage of the study?

We are interested in testing the hypothesis that the proportion of exposed individuals that are disease positive differs from the proportion of non-exposed individuals that are disease positive. Because this is nominal (count) data, a chi-squared test is the appropriate method to test this hypothesis. This involves three steps:

1. A statement of the null hypothesis: 'The proportion of exposed individuals that are diseased does not differ from the proportion of non-exposed individuals that are diseased.'

2. Calculation of a chi-squared test statistic. Using the standard notation the formula for the chi-squared test statistic for data presented in a 2 × 2 table is:

$$\chi_1^2 = \frac{n(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)} \tag{30}$$

3. We will use an alpha level of 0.05 to test this hypothesis and apply a one-tailed test. Specifying an alpha level of 0.05 means that there is a 5% probability of incorrectly rejecting the null hypothesis (when it is in fact true). The critical value that separates the upper 5% of the χ^2 distribution with 1 degree of freedom from the remaining 95% is 3.841 (from statistical tables). Thus, if our calculated chi-squared test statistic is greater than 3.841 we can reject the null hypothesis and accept the alternative hypothesis, concluding that the proportions diseased among exposed and non-exposed individuals differ.

11.6 Recommendations

What recommendations, if any, would you make to your colleague and to his client based on your findings (without the data from the clinical trial or cohort study)?

11.7 Clinical trial

Design either a clinical trial or a prospective cohort study to test one of your hypotheses in detail.

11.8 Financial impact

Estimate the financial impact of the losses due to diarrhoea in this set of 26 litters. The following data has been provided:

Litter	Pen	Sow	Parity	Farrow	Born	Weaned	Death due to		
							Overlay	Scours	Other
1	9	124	1	03 Apr 02	12	9	1	2	0
2	4	121	1	03 Apr 02	9	6	1	2	0
3	12	76	3	04 Apr 02	8	8	0	0	0
4	13	164	2	05 Apr 02	11	9	0	2	0
5	16	27	6	06 Apr 02	7	7	0	0	0
6	1	18	4	09 Apr 02	10	6	0	4	0
7 [a]	7	3	2	10 Apr 02	14	8	2	2	2
8	3	69	8	10 Apr 02	10	9	1	0	0
9	11	13	5	11 Apr 02	8	8	0	0	0
10	2	101	3	12 Apr 02	12	7	2	1	2
11	8	83	6	14 Apr 02	11	10	1	0	0
12	5	79	2	15 Apr 02	11	11	0	0	0
13	10	62	4	18 Apr 02	9	8	1	0	0
14 [a]	6	74	1	18 Apr 02	10	7	0	3	0
15	4	27	1	19 Apr 02	9	6	0	3	0
16	15	61	7	23 Apr 02	6	5	1	0	0
17	12	52	5	24 Apr 02	12	10	0	0	2
18	3	107	2	26 Apr 02	15	9	4	2	0
19	16	27	3	26 Apr 02	10	9	1	0	0
20	1	159	1	27 Apr 02	6	6	0	0	0
21	13	41	2	28 Apr 02	6	6	0	0	0
22	7	131	4	29 Apr 02	8	6	0	2	0
23	9	83	6	30 Apr 02	7	6	0	0	1
24	2	79	3	30 Apr 02	9	9	0	0	0
25	8	128	5	30 Apr 02	12	10	1	1	0
26	11	169	4	30 Apr 02	11	10	0	0	1
Total					253	205	16	24	8

[a] Sow sick at farrowing.

Item	Value	Target
Percent of litters with scours in 12 months before outbreak	7%	$< 5\%$
Preweaning mortality in 12 months before outbreak	11.5%	$< 12\%$
Post weaning mortality	5%	$< 3\%$
Gross margin per pig marketed	\$35.00	-
Treatment costs per litter	\$10.00	-
E. coli vaccine	$2 \times$ \$2.50	-
Labour cost to vaccinate one pig	\$0.30	-

12 Review questions

12.1 Host, agent, environment

You are discharging a 2 year-old male domestic shorthair cat who has spent 10 days in your clinic recovering from the complications associated with obstruction of the urinary tract. As the cat's owner is writing out a cheque for $1500 he asks 'will my cat experience another attack of FUS in the future and what can I do to prevent it?' What advise would you give, from an epidemiological perspective?

Think about three or four health problems or diseases that you or your friends have had. List each of the host, agent, and environmental factors that may have been causative for each disease you have listed.

Can you think if circumstances when exposure to a causal factor does not change disease incidence?

List five or six broad and fundamental influences on health and disease, that is, those influences that change the population patterns of disease.

Reflect on some medical and public health activities which were widely practiced but are now known to be wrong, some dangerously so. Your reflection should include some historical activities say, before the turn of the twentieth century and more recent ones. Also reflect on some current policies and practices that may meet the same fate.

12.2 Measures of health

Imagine you are in a country where no animal demographic data is available. An epidemic of pneumonia is suspected in the cattle population. You are asked to develop a plan to prevent and control the epidemic. Which questions do you need to answer to start a rational control strategy for this disease? Which epidemiological data do you need to answer the questions?

What benefits are there from investigating the changes in disease frequency in a population over time?

Consider the reasons why a variation in disease pattern might be artefact rather than real. Can you group them into three or four categories of explanation? What explanations can you think of for a real change in disease frequency? Can you group these into three or four categories of explanation?

Imagine you are asked to describe the health status of a population of animals to a senior public servant. The person you are talking to has no previous background in animal (or human) health. What kinds of measures would you choose to portray the health of the animal population? Consider not only the specific types of data, but also the qualities of the data you would seek out.

Imagine a population of 10,000 new army recruits. You are interested in studying the incidence and prevalence rate of gunshot wounds on war duty. Assume all gunshot wounds lead to permanent visible damage. You follow the recruits for one year. All of the study population survive, all medical records are available, and all recruits are available to interview and examination. Assume the occurrence of gunshot wounds is spread evenly through the year, and that at the

time of entering the army, no recruits had gunshot wounds. Over the year you determine that 20 recruits had a gunshot wound.

- What is the incidence risk of gunshot wounds? What is the incidence rate of gunshot wounds?
- What is the point prevalence rate of having had a gunshot wound at the beginning, middle, and end of the year?
- What is the period prevalence rate over the year?
- If the incidence rate remains the same over time, what is the prevalence rate of ever being scarred by the end of five years?
- What is the average duration of a gunshot wound, among those scarred, by the end of the first year?
- What is the estimated point prevalence rate over the five-year period?

What might be your denominator for a study defining the incidence rate of:

- Calf mortality.
- Clinical mastitis.
- Bovine spongiform encephalopathy.

12.3 Measures of association

Reflect on the terms 'risk factor' and 'cause of disease.' What is the difference between these terms?

Consider why the risk ratio might provide a false picture of the effect of a risk factor on disease and hence the strength of association.

Imagine that the incidence of chronic obstructive pulmonary disease (COPD) in horse is compared in two areas of a country: one with polluted air (A) and the other not (B). In the polluted area there were 20 cases of COPD in a population of 100,000. In the other area there were 10 cases in a population of 100,000.

- What is the risk ratio of COPD in area (A)?
- What is the risk ratio of COPD in area (B)?
- Do we know the precision of these estimates of risk ratio?
- What explanations are there for the risk ratio estimate in area (A)?
- What questions will you need to consider before concluding that there is a real association between pollution and COPD?

Imagine that exposure to a dry cat food triples the incidence of a feline urologic syndrome (FUS), that is, the risk ratio is 3. This disease has a baseline incidence of 1 per cent per year in the non-exposed group. Imagine also that the baseline incidence is double in castrated male cats (that is, 2 per cent) and that the risk ratio associated with exposure to dry cat food is the same, three. You follow 100 entire and 100 castrated male cats that are fed dry cat food, and an equivalent number of cats fed moist food. The study lasts for 5 years. Create a 2×2 table to show the data for castrates and entire male cats and calculate the odds ratio of disease in the exposed group in relation to those not exposed. Compare the odds ratio with the risk ratio of 3.

The Ministry of Health has made available a sum of $100,000 for a health promotion programme to reduce coronary heart disease mortality. We can spend it on encouraging people to stop smoking or encouraging them to do more exercise. Assume the risk ratio associated with both risk factors is 2, that changes in prevalence rate are equally permanent, and that the cardioprotective effect occurs quickly. Which choice will give a better return in lives saved?

- First, make a judgement on which of the two preventive programs you prefer.
- Now consider which is more common: smoking or lack of exercise?
- Calculate the population attributable risk when the prevalence rate of smoking is 20%, 30%, 40% and 50% and the prevalence rate of lack of exercise is 60%, 70%, and 80% (these are realistic prevalence rates in industrialised countries). Has the result altered or substantiated your earlier judgement?

12.4 Study design

Imagine a cohort study which aims to determine the incidence of arthritis in large breeds of dogs. The follow-up period for the study is five years. Describe the advantages and disadvantages of the two approaches for measuring incidence.

Imagine a study of the incidence of congestive heart disease in large breeds of dogs, base on post mortem records collected at a University teaching hospital over a five-year period. Again, consider the advantages and disadvantages of the two approaches for measuring incidence.

Is there a difference between a clinical case series and a population case series?

How might epidemiology study the potential role in disease causation of factors which vary little between individuals within a region or country. For example: fluoride content of water, hardness or softness of water supplies, annual exposure to sunshine?

What is the essential feature that differentiates a cross-sectional study from a cohort study?

Explain what you understand by the term ‘error’. What is the difference, if any, between error and bias?

12.5 Diagnostic tests

A client of your manages a study beef herd which, for the past ten years, has consistently tested negative for tuberculosis. A positive reactor has been found after the latest round of testing. What would you advise?

13 Resources

EpiCentre, Massey University	http://epicentre.massey.ac.nz/
University of Guelph, Department of Pop Medicine	http://www.ovc.uoguelph.ca/popm/
Royal Veterinary College, University of London	http://www.rvc.ac.uk/
University of Michigan School of Public Health	http://www.sph.umich.edu/epid/
Epidemiology Monitor	http://www.epimonitor.net/
Association of Teachers of Veterinary Public Health	http://www.cvm.uiuc.edu/atvphpm/
Epidemiology for the uninitiated — BMJ	http://www.bmj.com/epidem/
Centers for Disease Control and Prevention	http://www.cdc.gov/
EXCITE	http://www.cdc.gov/excite/
Epidemiology Supercourse	http://www.pitt.edu/~super1/
VEIN links: Evidence Based Medicine	http://vein.library.usyd.edu.au
Post Graduate Foundation in Veterinary Science	http://www.pgf.edu.au/
EBM Resources	http://www.dartmouth.edu/~library/biomed/
MAF, New Zealand	http://www.maf.govt.nz
DAFF, Australia	http://www.daff.gov.au/
Canadian Food Inspection Agency	http://www.inspection.gc.ca
Health Canada	http://www.hc-sc.gc.ca/
Instituto Nacional de Tecnología Agropecuaria	http://www.inta.gov.ar/
International EpiLab	http://www.food.dtu.dk/Default.aspx?ID=9406
The Cochrane Collaboration	http://www.cochrane.org

References

Altman, D., & Bland, J.(1994a). Statistics Notes: Diagnostic tests 1: sensitivity and specificity. *British Medical Journal*, *308*, 1552.

Altman, D., & Bland, J.(1994b). Statistics notes: Diagnostic tests 2: predictive values. *British Medical Journal*, *309*, 102.

Anderson, R., Crespo, C., Bartlett, S., Cheskin, L., & Pratt, M.(1998). Relationship of physical activity and television watching with body weight and level of fatness among children. Results from the Third National Health and Nutrition Examination Survey. *Journal of the American Medical Association*, *279*, 938 - 942.

Ast, D., & Schlesinger, E. (1956). The conclusion of a ten-year study of water fluoridation. *American Journal of Public Health*, *46*, 265 - 271.

Brenner, H., Greenland, S., & Savitz, D. (1992). The effects of nondifferential confounder misclassification in ecological studies. *Epidemiology*, *3*, 456 - 469.

Carey, J., Klebanoff, M., Hauth, J., Hillier, S., Thom, E., & Ernest, J. (2000). Metronidazole to prevent preterm delivery in pregnant women with asymptomatic bacterial vaginosis. *New England Journal of Medicine*, *342*, 534 - 540.

Dawson-Saunders, B., & Trapp, R.(1994). *Basic and Clinical Biostatistics*. New York: Appleton and Lange.

Deeks, J., & Altman, D.(2004). Statistics Notes: Diagnostic tests 4: likelihood ratios. *British Medical Journal*, *329*, 168 - 169.

Dohoo, I., Martin, S., & Stryhn, H.(2003). *Veterinary Epidemiologic Research*. Charlottetown, Prince Edward Island, Canada: AVC Inc.

Donnelly, C., Ghani, A., Leung, G., Hedley, A., Fraser, C., Riley, S., et al.(2004). Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet*, *361*, 1761 - 1766.

Draper, G., Vincent, T., Kroll, M., & Swanson, J. (2005). Childhood cancer in relation to distance from high voltage power lines in England and Wales: a case-control study. *British Medical Journal*, *330*, 1290.

Elwood, J.(1988). *Causal Relationships in Medicine*. New York, USA: Oxford University Press.

Elwood, J.(1998). *Critical Appraisal of Epidemiological Studies and Clinical Trials*. New York, USA: Oxford University Press.

Farquharson, B.(1990). On-farm trial. In D. Kennedy (Ed.), *Epidemiological Skills in Animal Health. Refresher Course for Veterinarians. Proceedings 143* (p. 207 - 212). Postgraduate Committee in Veterinary Science, University of Sydney, Sydney, Australia.

Fletcher, R., Fletcher, S., & Wagner, E.(1996). *Clinical Epidemiology*. Baltimore, USA: Williams and Wilkins.

Fransen, M., Woodward, M., Norton, R., Robinson, E., Butler, J., & Campbell, A. (2002). Excess mortality or institutionalisation following hip fracture: men are at greater risk than women. *Journal of the American Geriatrics Society*, *50*, 685 - 690.

Fraser, D., Tsai, T., Orenstein, W., Parkin, W., Beecham, H., Sharrar, R., et al.(1977). Legionnaires' disease — description of an epidemic of pneumonia. *New England Journal of Medicine*, *296*, 1189-1197.

Gardner, I. (1990a). Case study: Investigating neo-natal diarrhoea. In D. Kennedy (Ed.), *Epidemiology at Work. Refresher Course for Veterinarians. Proceedings 144* (p. 109 - 129). Postgraduate Committee in Veterinary Science, University of Sydney, Sydney, Australia.

Gardner, I.(1990b). Reporting disease outbreaks. In *Epidemiology at Work: Refresher Course for Veterinarians* (p. 29 - 42). Quarantine Station, North Head, NSW: Post Graduate Committee in Veterinary Science, The University of Sydney.

Gardner, L., Landsittel, D., & Nelson, N.(1999). Risk factors for back injury in 31,076 retail merchandise store workers. *American Journal of Epidemiology*, *150*, 825 - 833.

Goodwin-Ray, K., Stevenson, M., & Heuer, C.(2008c). Flock-level casecontrol study of slaughter-lamb pneumonia in New Zealand. *Preventive Veterinary Medicine*, *85*, 136 - 149.

Greenhalgh, T.(1997a). How to read a paper: Assessing the methodological quality of published papers. *British Medical Journal*, *315*, 305 - 308.

Greenhalgh, T.(1997b). How to read a paper: Getting your bearings (deciding what the paper is about). *British Medical Journal*, *315*, 243 - 246.

Greenhalgh, T.(1997c). How to read a paper: Papers that report diagnostic or screening tests. *British Medical Journal*, *315*, 540 - 543.

Greenhalgh, T.(1997d). How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). *British Medical Journal*, *315*, 672 - 675.

Greenhalgh, T.(1997e). How to read a paper: Papers that tell you what things cost (economic analyses). *British Medical Journal*, *315*, 596 - 599.

Greenhalgh, T.(1997f). How to read a paper: Statistics for the nonstatistician. II: 'Significant' relations and their pitfalls. *British Medical Journal*, *315*, 422 - 425.

Greenhalgh, T.(1997g). How to read a paper: The Medline database. *British Medical Journal*, *315*, 180 - 183.

Greenhalgh, T.(2006). *How to Read a Paper: The Basics of Evidence-Based Medicine*. London: British Medical Journal Books.

Greenhalgh, T., & Taylor, R.(1997). How to read a paper: Papers that go beyond numbers (qualitative research). *British Medical Journal*, *315*, 740 - 743.

Hill, A.(1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of London. Series C, Medicine*, *58*, 295 - 300.

Hoyert, D., Arias, E., Smith, B., Murphy, S., & Kochanek, K.(1999). *Deaths: final data for 1999. National Vital Statistics Reports Volume 49, Number 8.* Hyattsville MD: National Center for Health Statistics.

Johansen, C., Boise, J., McLaughlin, J., & Olsen, J.(2001). Cellular telephones and cancer — a nationwide cohort study in Denmark. *Journal of the National Cancer Institute*, *93*, 203 - 237.

Kelsey, J., Thompson, W., & Evans, A.(1986). *Methods in Observational Epidemiology.* London: Oxford University Press.

Leung, W.-C.(2002). Measuring chances. *Student British Medical Journal*, *10*, 268 - 270.

Levy, P., & Lemeshow, S.(1999). *Sampling of Populations Methods and Applications.* London: Wiley Series in Probability and Statistics.

Mackintosh, C., Schollum, L., Harris, R., Blackmore, D., Willis, A., Cook, N., et al. (1980). Epidemiology of leptospirosis in dairy farm workers in the Manawatu. Part I: A cross-sectional serological survey and associated occupational factors. *New Zealand Veterinary Journal*, *28*, 245 - 250.

Martin, S., Meek, A., & Willeberg, P.(1987). *Veterinary Epidemiology Principles and Methods.* Ames, Iowa: Iowa State University Press.

Morris, R.(1990). Disease outbreak! What can you do? In D. Kennedy (Ed.), *Epidemiological Skills in Animal Health. Refresher Course for Veterinarians. Proceedings 143* (p. 321 - 327). Postgraduate Committee in Veterinary Science, University of Sydney, Sydney, Australia.

Muscat, J., Malkin, M., Thompson, S., Shore, R., Stellman, S., & McRee, D.(2000). Handheld cellular telephone use and risk of brain cancer. *Journal of the American Medical Association*, *284*, 3001 - 3007.

Noordhuizen, J., Frankena, K., Hoofd, C. van der, & Graat, E.(1997). *Application of Quantitative Methods in Veterinary Epidemiology.* Wageningen: Wageningen Pers.

Oleckno, W. (2002). *Essential Epidemiology Principles and Applications.* Prospect Heights, Illinois: Waveland Press.

Olsen, J.(2003). What characterises a useful concept of causation in epidemiology? *Journal of Epidemiology and Community Health*, *57*, 86 - 88.

Parsonnet, J., Friedman, G., Vandersteen, D., Chang, Y., Vogelman, J., Orentreich, N., et al. (1991). *Helicobacter pylori* infection and the risk of gastric-carcinoma. *New England Journal of Medicine*, *325*(16), 1127 - 1131.

Petrie, A., & Watson, P.(2005). *Statistics for Veterinary and Animal Science.* London: Blackwell Science.

Pfeiffer, D., Robinson, T., Stevenson, M., Stevens, K., Rogers, D., & Clements, A.(2008). *Spatial Analysis in Epidemiology.* New York, USA: Oxford University Press.

Rinzin, K., Stevenson, M., Probert, D., Bird, R., Jackson, R., French, N., et al. (2008). Free-roaming and surrendered dogs and cats submitted to a humane shelter in Wellington, New Zealand, 1999 – 2006. *New Zealand Veterinary Journal*, *In press*.

Rogan, W., & Gladen, B.(1978). Estimating prevalence from results of a screening test. *American Journal of Epidemiology*, *107*, 71 - 76.

Rothman, K.(1976). Causes. *American Journal of Epidemiology*, *104*, 587 - 592.

Rothman, K., & Greenland, S.(1998). *Modern Epidemiology*. Philadelphia, USA: Lippincott - Raven.

Schlesselman, J. (1982). *Case-Control Studies Design, Conduct, Analysis*. London: Oxford University Press.

Schwarz, D., Grisso, J., Miles, C., Holmes, J., Wishner, A., & Sutton, R.(1994). A longitudinal study of injury morbidity in an African-American population. *Journal of the American Medical Association*, *271*, 755 - 760.

Selvin, S.(1996). *Statistical Analysis of Epidemiological Data*. London: Oxford University Press.

Siscovick, D., Weiss, N., Fletcher, R., & Lasky, T.(1984). The incidence of primary cardiac-arrest during vigorous exercise. *New England Journal Of Medicine*, *311*, 874 - 877.

Smith, R. (1995). *Veterinary Clinical Epidemiology — A Problem-Oriented Approach*. Boca Raton, Florida: CRC Press.

Stevenson, M., Morris, R., Lawson, A., Wilesmith, J., Ryan, J., & Jackson, R.(2005). Area-level risks for BSE in British cattle before and after the July 1988 meat and bone meal feed ban. *Preventive Veterinary Medicine*, *69*, 129 - 144.

Stevenson, M., Wilesmith, J., Ryan, J., Morris, R., Lawson, A., Pfeiffer, D., et al. (2000). Descriptive spatial analysis of the epidemic of bovine spongiform encephalopathy in Great Britain to June 1997. *Veterinary Record*, *147*, 379 - 384.

Thrusfield, M.(2005). *Veterinary Epidemiology*. London: Blackwell Science.

Trivier, J., Caron, J., Mahieu, M., Cambier, N., & Rose, C. (2001). Fatal aplastic anaemia associated with clopidogrel. *Lancet*, *357*, 446.

Valent, F., Brusaferro, S., & Barbone, F.(2001). A case-crossover study of sleep and childhood injury. *Pediatrics*, *107*, E23.

Vander Stoep, A., Beresford, S., & Weis, N.(1999). A didactic device for teaching epidemiology students how to anticipate the effect of a third factor on an exposure-outcome relation. *American Journal of Epidemiology*, *150*, 221.

Webb, P., Bain, C., & Pirozzo, S.(2005). *Essential Epidemiology*. Cambridge, UK: Cambridge University Press.

Wilesmith, J., Stevenson, M., King, C., & Morris, R. (2003). Spatio-temporal epidemiology of foot-and-mouth disease in two counties of Great Britain in 2001. *Preventive Veterinary Medicine*, *61*, 157 - 170.

Will, R., Ironside, J., Zeidler, M., Cousens, S., Estibeiro, K., & Alperovitch, A. (1996). A new variant of Creutzfeld-Jacob disease in the UK. *Lancet*, *347*, 921 - 925.

Yang, C., Chiu, H., Cheng, M., & Tsai, S. (1998). Chlorination of drinking water and cancer in Taiwan. *Environmental Research*, *78*, 1 - 6.