Disease does not occur at random; if it were we would not have a job! There is a pattern for every disease; we just need to find it. To find how disease behaves we need to answer the following questions:

- What is the problem?
- Who gets diseased?
- Where is the disease concentrated?
- When does disease occur?

Answering all these questions (the essence of epidemiology is describing disease in populations) should lead us to the answer of the ultimate question we have about a certain disease (why does it happen?) and enable us to prevent it.

**Case definition**

The best explanation of the true substance of the word “definition” in matters pertinent to epidemiology comes from combining two of the meanings of the “definition”: (i) an exact statement or description of the nature, scope, or meaning of something, and (ii) the degree of distinctness in outline of an object (Oxford Dictionaries online). Therefore, the more carefully we describe things, the more distinctness we achieve. In defining words, it is important to avoid using another word with the same root as the one we are defining. When defining a case, it tends to be more complete and accurate when following the same rule of not using words with the same root.
What is the problem?

Before we start looking into who is diseased or where it is, we need to define what we are going to consider a diseased individual looks like; in other words, we need a **case definition**. This seems silly at first, but it is the most important step in any study or investigation and is not so clear-cut if you look deeper.

**Example**

When asked to define a diarrheic patient, simply stating it is a dog with diarrhea does not give much distinction to the case. However, if we define a diarrheic patient as a dog with feces that are not well-formed and cannot be picked up without leaving a mark on the ground gives a clear-cut characteristic that allows anyone to categorize a patient as having diarrhea or not.

The importance of case definition becomes paramount when comparing research studies about a certain disease. If two studies do not have the same case definition, the results of both studies cannot be compared directly.

**Example**

Let us suppose we want to investigate if there is a problem of parvovirus in a kennel. How would you define a case of parvovirus? Most people would say a puppy with diarrhea. The problems with this simple definition of a case of parvovirus are as follows:

- There are other causes of diarrhea in puppies, so you may be overestimating how much parvovirus infection there truly is.
- Parvovirus may have asymptomatic infections, so you may be underestimating infection.
- Parvovirus can have other clinical signs without diarrhea, such as lethargy, anorexia, fever, vomiting, and severe weight loss, so you may be underestimating infection by looking only at puppies with diarrhea.
- How old can a dog be while still being considered a puppy? In other words, what is the “case definition” of a puppy?

To get the best estimate of truly infected dogs in a population, we would have to better define a case of parvovirus infection. An example could be “dogs less than 9 months old with a positive fecal ELISA test for parvovirus.” This definition would minimize the number of dogs with diarrhea due to other causes (because they have to have a positive ELISA test), and it would also minimize the number of dogs excluded because they did not have diarrhea.

Example

A study on hip dysplasia in dogs (Paster et al. 2005) showed that inclusion of the caudal curvilinear osteophyte in the definition of canine hip dysplasia significantly altered the diagnosis of a large proportion of dogs, usually toward a higher score but sometimes to a lower score (Figure 1.1).
Another example is from a study on diagnosis of staphylococcal infections in a veterinary hospital (Geraghty et al. 2013). In this study, phenotypic appearance of cultured bacteria or genotypic analysis was used to determine which staphylococcal species was isolated from each animal. Figure 1.2 shows a summary of the data presented in the published paper, showing large mismatch in the results using one method versus the other.


**Figure 1.2** Distribution of isolation of staphylococcal species defined via phenotypic or genotypic methods (data source Geraghty, L., Booth, M., Rowan, N., and Fogarty, A. (2013). Investigations on the efficacy of routinely used phenotypic methods compared to genotypic approaches for the identification of staphylococcal species isolated from companion animals in Irish veterinary hospitals. *Irish Veterinary Journal*, 66(1):7–15).
Case definition is of paramount importance in situations where a range of outcomes is possible. This is typical of outcomes that are measured by scores, which are used to establish a relative degree of the outcome when there is no directly measurable factor.

**Example**

In a study on gastric ulcers in pleasure horses (Niedzwiedz et al. 2013), the authors used a scoring system to determine the severity of the lesion. The scoring system they described is shown in Figure 1.3. Notice that with this description it would be possible to replicate the study using the same scoring system and therefore comparing results across studies. There could be only a potential problem in determining what “small” and what “large” lesions are—that is, a diameter threshold that would qualify a lesion as small or large. Therefore, it is better to always use objective characteristics to define cases or scores.

<table>
<thead>
<tr>
<th>Lesion severity score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No lesions</td>
</tr>
<tr>
<td>I</td>
<td>Lesions appear superficial (only mucosa missing)</td>
</tr>
<tr>
<td>II</td>
<td>Small, single, or multifocal erosions or ulcers</td>
</tr>
<tr>
<td>III</td>
<td>Large, single, or multifocal ulcers, or extensive erosions and sloughing</td>
</tr>
<tr>
<td>IV</td>
<td>Active hemorrhage or adherent blood clot</td>
</tr>
</tbody>
</table>

**Figure 1.3** Lesion severity score description for a study on gastric lesions in pleasure horses (Niedzwiedz, A., Kubiak, K., & Nicpon, J. (2013). Endoscopic findings of the stomach in pleasure horses in Poland. *Acta Veterinaria Scandinavica, 55*:45–55).

**Who is affected?**

Remember we are looking for patterns of disease, so the question is whether the entire population is affected or there are some specific **subgroups** more affected than others? Any type of subgrouping can be investigated: age, gender, breed, environment, disposition (mainly used for companionship, racing, hunting, or other), diet, etc. To continue with the parvovirus example, we know that most affected animals are puppies and young dogs. Among the young dogs it is mostly males, in theory reflecting their higher tendency to roam lose compared with females.

An example for the environmental differences can be found in feline leukemia, a disease more common in multicat households and in cats that are allowed access to the outdoors.

You can surely find an example for different diets, breeds, etc.
Where is the disease concentrated?
Defining the spatial distribution of disease may help identify risk factors and the behavior of infection. A risk factor is any characteristic that increases the risk of an animal for a certain condition. For example, which horses get infected, those in pasture or those in the barn? Is the disease spreading to adjacent stalls or are apparently “random” stalls involved? Are neighboring farms affected too? Do affected animals live in specific areas such as downtown (smog), or close to wet areas?

When does disease occur?
Is there a pattern in time? How many animals are affected in winter versus summer, spring, and fall? Is there a difference in the number of diseased individuals before and after a given event (change in disinfectant, vaccination event, etc.)? Is there a cyclical nature to the disease that could coincide with mosquito season or freezing?

Evaluate the epidemic curve—temporal distribution of cases. The first case diagnosed in an outbreak is called the “index case.” A representation of the number of cases by days will show the type of epidemic curve of a disease (Figure 1.4). A “point-source” curve shows a high number of affected animals initially, which fades over time. This is typical of situations where many animals are exposed at the same time, like in outbreaks of food-borne diseases. A “propagated” epidemic curve shows a slow increase in the number of cases and a slow decrease too. This curve is typical of epidemics of infectious (contagious) diseases, where animals get exposed at different points in time (i.e., one animal gets infected and spreads the infection to a few others, which in turn infect others).

Types of measurements
Following are the most common ways to measure events in epidemiology, and then we will look into specific measurements of disease.

Counts
A count of individuals is used to establish the size of the population. However, when evaluating how important a disease is, simply reporting the count of sick animals does not give much useful information.

Example
If someone says they have two sick dogs, is that a little or a lot? Obviously, it depends on how many dogs they have in total. If they have two dogs, it means all of their dogs are diseased, but if it is a kennel that has 50 dogs, 2 out of 50 dogs is not a lot.
Everything has to be studied in context, in the case of epidemiology, in reference of the total population. Some may be thinking now that if we are dealing with a terrible disease that can spread very fast and kill the animals, even 2 out of 50 animals is too much. Agreed, but it is not a lot compared with 2 out of 2. We are simply looking at numbers right now; we will add meaning or significance to these numbers later in Chapter 5. The point is that, to give a sense of

**Figure 1.4** Epidemic curves: point-source (top) and propagated (bottom).
how big the number of diseased animals is, it needs to be put in context in reference of the size of the total population.

**Proportions**

A proportion is the most normal way of looking at the magnitude of the number of animals affected with a disease. It puts the count of sick animals in perspective of the number of total animals in the population.

The formula to calculate a proportion is as follows:

\[
\frac{A}{A + B}
\] (1.1)

where \(A\) is the number of sick animals and \(B\) is the number of healthy animals. Together \(A\) and \(B\) make the total population.

Note that the numerator is ALWAYS included in the denominator. Therefore, *proportions compare a subgroup with the whole group of animals under study*. They are usually expressed as percentages.

**Example**

Two sick dogs would represent 100% for the client that has two dogs total:

\[
\frac{\text{Sick}}{\text{Sick + Healthy}} = \frac{2}{2 + 0} = 1 = 100\%
\]

While in a kennel that has 50 dogs, they would represent only 4%:

\[
\frac{\text{Sick}}{\text{Sick + Healthy}} = \frac{2}{2 + 48} = 0.04 = 4\%
\]

When calculating and reporting proportions, it is paramount to report what population is included in the denominator, as this may not always be clear, and simply reporting a percentage can lead to confusion as to how that proportion was calculated.

**Example**

In a study about risk factors for dystocia in Boxers (Linde Forsberg and Persson 2007), the authors show a graph (Figure 1.5) with two different proportions calculated using the same animals in the numerator but different denominator. The light bars represent the proportion of bitches within each age group (numerator) among all whelpings (denominator, \(n = 253\)), while the dark bars represent the proportion of bitches within each age group (numerator) among whelpings that resulted in dystocia (denominator, \(n = 70\)). This is not clear from the graph itself but becomes evident when reading the text.
Ratios

A ratio shows the relationship between two mutually exclusive groups. This means that the numerator cannot be included in the denominator. In other words, an animal cannot be part of both groups that are being compared. It is like comparing apples and oranges.

The formula to calculate a ratio is as follows:

\[ \frac{A}{B} \]
where \( A \) is the number of animals in one group and \( B \) is the number of animals in the other group.

A typical example of a ratio you can see in the literature is the ratio of males to females. Obviously, an animal cannot be both. It is usually expressed in print with figures as \( A:B \) and with text as \( A/B \) or \( A\text{-to-}B \). Verbally, it is expressed as “ratio of \( A \) to \( B \).” It does not matter which one of the two groups goes first, although there seems to be a tendency to put the lowest number last.

### Example

A typical veterinary clinic may be expected to have a 5:1 dog-to-cat visits. This means that for each cat they see, the clinic will see five dogs. Again, it is obvious that an animal cannot be both a dog and a cat, so this is a ratio.

In another example, it has been shown that a higher adult/young ratio decreases aggression among young horses. This means that the more adult horses there are for each young horse, the better they all get along. Horses are either young or old; they cannot be both at the same time.

However, it is not always easy to determine where to draw the line to include an animal into one group or another when the characteristic that is used to classify them changes over time, as opposed to gender or breed, which are fixed. With the example of the horses, we could consider that a horse is young until 3 years of age. So a horse that is 2 years and 11 months old (35 months) will be considered “young,” while a horse that is 3 years and 1 month old (37 months) will be considered old. Do we really expect much difference in behavior between these two horses? Should they be included when studying horse aggression? Should we use a different cutoff point for this study? These are some of the most common questions that arise when dealing with ratios. Notice the importance of definitions of age in this case.

### Rates

A **rate** represents the speed of something developing. A rate compares a subgroup with the whole group of animals during a specific time. Therefore, it is like looking at a proportion *including the time each individual is at risk*.

The formula to calculate a rate is as follows:

\[
\frac{A}{(A + B) \text{ time}}
\]  \hspace{1cm} (1.3)

The most important feature of a rate, which makes it different from a proportion, is that it directly accounts for the time that each individual is at risk.
Rates are very important when dealing with dynamic populations where animals come and go as part of the population. You are probably thinking right now that this is practically everywhere you work: your clinic, a kennel, the local shelter, a horse track, etc., and you are right. This is why epidemiology is so important to the clinical veterinarian, and why it is important to understand this measurement well. Any time when two animals are exposed unequal times to a potential risk factor for disease, we need to take those differences in “time at risk” into account.

### Specific measurements of disease

There are some specific measurements of disease that are commonly used in epidemiology, giving us information about how important (quantitatively) a disease is in a given population. There are two main measurements of disease: prevalence and incidence.

#### Prevalence

Prevalence is a proportion that describes the number of animals that have a certain condition of interest at a given time. The formula to calculate prevalence has the number of animals that have that condition during the time of study in
the numerator, divided by the number of animals at risk of developing the condition that are present at during that same time (denominator).

The formula to calculate prevalence is as follows:

\[
\frac{\text{Total no. of cases}}{\text{Population at risk}}
\] (1.4)

Because prevalence is a proportion, it is expressed as a percentage.

**Example**

Assume that in the past year you have seen 700 canine patients in your clinic, 120 of which were new puppies for their vaccinations. They all received three doses of canine distemper vaccines according to label (3–4 weeks apart before 16 weeks of age). In spite of this, 3 puppies developed signs of distemper. The prevalence of distemper among puppies in your clinic last year was \(\frac{3}{120} = 2.5\%\).

Only the population at risk should be included in the denominator, that is, animals that can experience the event in the numerator. In the example aforementioned, only puppies are included, not all dogs. Other examples of accurately selecting the denominator for the calculation of prevalence would be including only intact males in the denominator for calculating the prevalence of testicular cancer or including only pregnant females when evaluating the prevalence of abortions (only pregnant females can abort). This is not complicated but requires some attention.

**Example A**

Figure 1.7 can represent both cats at the local shelter or horses at a racetrack, whatever you prefer. Each line represents a different animal identified by name. Each column represents 1 week. The gray horizontal bars represent the presence of the animal on the premises, while each triangle represents a case of respiratory disease. Black triangles represent the first time the animal shows respiratory signs, while white triangles represent recurring cases.

The prevalence of respiratory disease during the 12-week period is as follows:

- **Numerator:** total cases of respiratory disease = 6 new + 2 recurring = 8 (count all triangles)
- **Denominator:** number of animals on the premises at any time during the period in question = 15 (count horizontal bars)

\[
\text{Prevalence} = \frac{8}{15} = 0.533 = 53.3\%
\]

Prevalence is expressed as a percentage; therefore, the prevalence of respiratory disease in these facilities was 53.3% during the 12-week period.
Example B

Now let us assume that we are only interested in the first 4 weeks of this period (Figure 1.8). The adjusted chart would look like this:

The prevalence of respiratory disease during this 4-week period is as follows:

Figure 1.7 Graphic representation of prevalence calculation (Example A).

Figure 1.8 Graphic representation of prevalence calculation (Example B).
Incidence

Incidence is a rate that describes the speed at which a given population acquires or develops a certain condition. To calculate the incidence, only the number of new cases that occurred during the evaluated period of time is included in the numerator, while the denominator takes into account the time that each animal is at risk. This is important because once an animal has acquired a certain condition (e.g., been neutered, aborted, or developed diabetes), it is not at risk of “newly” developing that condition again, at least within a certain period of time. For example, a female can abort multiple times but only when she is pregnant.

The formula to calculate incidence is as follows:

\[
\text{Incidence} = \frac{\text{No. of new cases}}{\text{Population-time at risk}}
\]

(1.5)

Because incidence is a rate, it has to be expressed using the appropriate time units (cat-days, horse-weeks, etc.). Commonly, the reporting is done in whole integers (without decimals), although it is not compulsory. In other words, an incidence of 0.25 cases per cow-day would commonly be reported as 25 cases per 100 cow-days.

<table>
<thead>
<tr>
<th>Example C</th>
</tr>
</thead>
</table>

Let us go back to the example of cats in a local shelter or horses at a racetrack (Figure 1.7). The incidence of respiratory disease during the entire 12-week period is as follows:

- **Numerator**: only new cases of respiratory disease = 6 (count only black triangles)
- **Denominator**: total weeks at risk up to when an animal has its first case (count individual dark gray cells). This can be easily visualized by changing the color of the weeks once an animal has suffered a case of respiratory disease, as seen in Figure 1.9. We count only the dark gray cells (Figure 1.9).

There are a total of 48 cat-weeks or horse-weeks of exposure. Therefore, the incidence of respiratory disease in these facilities is as follows:

\[
\text{Incidence} = \frac{6}{48} = 0.125
\]

Expressed as 0.125 cases per cat-week (or horse-week) or as 125 cases per 1000 cat-weeks (or horse-weeks).
Example D

If we were to look at the first 4 weeks only, the formula for calculating incidence would change to the following:
- **Numerator**: only new cases of respiratory disease = 5 (count only black triangles)
- **Denominator**: total weeks at risk up to when an animal has its first case (count individual dark gray cells) = 21

\[
\text{Incidence} = \frac{5}{21} = 0.238
\]

This is expressed as 0.238 cases per cat-week (or horse-week) or 238 cases per 1000 cat-weeks (or horse-weeks). The adjusted chart would look as in Figure 1.10.

**Animal name**

Skywalker
Alexa
Tahoe
Waylon
Xena
Mystic
Shamara
Midnight
Silver
Dante
Salem
Dynamite
Marengo
Thunder
Saab

**Figure 1.9** Graphic representation of incidence calculation (Example C, notice the altered colors).

**Figure 1.10** Graphic representation of incidence calculation (Example D).
Comparison of prevalence and incidence
The main differences between prevalence and incidence are as follows:

- **Prevalence** counts all cases in the population, while **incidence** only counts new cases.
- **Incidence** accounts for differences in time that animals are exposed to the risk of disease.

Prevalence can be compared with a photo of an event, while incidence would be the movie.

Therefore, both the numerator and the denominator can be different when calculating prevalence and incidence in a population. The numerator will be different if there are repeated cases of disease. The denominator will include time and will be different with repeated cases of disease. The denominator would also be different in dynamic populations (varying numbers of animals at risk). Once an animal has contracted a specific condition, it may not be at risk of developing the same condition again as a “new” event, although it can be a recurrence or recrudescence of the condition. Because it cannot be considered a “new” case, it is excluded from further incidence calculations. Hopefully, the example mentioned will help understand these subtleties.

**Example**

In the example of cats in a shelter or horses at a racetrack, we can see that although prevalence did not change much comparing the entire 12-week period and the initial 4 weeks, incidence was almost double in the initial 4-week period compared with the entire period, which indicates that the speed of disease was faster at the beginning of the period than at the end. The movie always gives you a better idea of what is going on than a single still photo.

Because of the difference in the numerator between prevalence and incidence, it is of utmost importance to properly define a new case, with special attention to the “new” part.

**Example**

If the same graphs were to represent lameness cases and lameness on different legs are considered different cases, the cases represented by the white triangles could in fact now be new cases if the lameness in that animal is on a different leg (so we would represent them as black triangles for ease of visualization).

The following last few measurements of disease are not as commonly reported in the veterinary literature but are presented here for ease of reference when they are encountered.
Morbidity

Morbidity is a very specific measurement of disease defined as the proportion of animals affected with a specific condition in a given population. Thus, it is a proportion. It is a measure of the amount of disease in a population, like prevalence.

The formula to calculate morbidity is as follows:

\[
\text{Morbidity} = \frac{\text{No. of cases}}{\text{Total population}}
\]  

(1.6)

Example

Assume a total population of 1000 dogs (all ages) that are seen by a veterinary clinic. Assume they see 6 dogs with gastric dilation/volvulus (GDV).

\[
\text{Morbidity of GDV is } \frac{6}{1000} = 6\%
\]

Mortality

Mortality is another specific measurement of disease defined as the number of animals that die of any cause within a population in a specific period of time. Thus, it is a rate and needs to include the time period in the denominator. It is also commonly referred to as crude mortality, to differentiate it from disease-specific mortality.

The formula to calculate mortality is as follows:

\[
\text{Mortality} = \frac{\text{Total no. of deaths}}{\text{Total population-time at risk}}
\]  

(1.7)

Example

Assume that from the total population of 1000 dogs seen by the veterinary clinic in the previous example, they lose 10 patients every month. For ease of calculation, we will focus on a single month.

Crude mortality in the population is 10/1000 dog-months = 0.01 deaths per dog-month.

Disease-specific mortality

This is another specific measurement used in epidemiology defined as the number of animals that die of a specific disease within a population in a specific period of time. Because it refers to mortality, it is also a rate. It is an indication as
to how many animals in a population die of a specific disease. It should not be confused with case-fatality, explained next.

The formula for disease-specific mortality is as follows:

$$\text{Disease-specific mortality} = \frac{\text{No. of deaths due to the disease}}{\text{Total population-time at risk}} \quad (1.8)$$

**Example**

Following with the previous example, assume that 2 of the 6 cases of GDV die in spite of everything they do to help them.

GDV-specific mortality is $$\frac{2}{1000}$$ dog-months = 0.002 deaths due to GDV per dog-month

**Case-fatality**

This measurement represents the severity of a disease. It is the proportion of diseased animals (denominator) that died due to the disease (numerator).

The formula to calculate case-fatality is as follows:

$$\text{Case-fatality} = \frac{\text{No. of deaths due to the disease}}{\text{No. of cases}} \quad (1.9)$$

**Example**

Using the numbers from the ongoing example,

GDV case-fatality is $$\frac{2}{6} = 33\%$$

The major difference between these four measurements can be more easily understood when expressing the outcome in a full sentence:

1. GDV morbidity: 6% of dogs seen at this clinic suffer from GDV.
2. Crude mortality: 1 dog dies every 100 dog-months (for comparison with the next measurement, we can express it as 10 dogs die every 1000 dog-months).
3. GDV-specific mortality: 2 dogs die of GDV every 1000 dog-months.
4. GDV case-fatality: 33% of dogs that have GDV die.

These four measurements are more easily visualized through a Venn diagram as follows (Figure 1.11).
1 Morbidity would be represented as the circle divided by the rectangle.
2 Mortality: the triangle divided by the rectangle.
3 Disease-specific mortality: intersecting slice of the triangle and the circle, divided by the rectangle.
4 Case-fatality: intersecting slice of the triangle and the circle, divided by the circle.

**Note**