# Module 2: Describing disease occurrence in populations

Contributors: Scott Wells, Amy Kinsley, Julio Alvarez

In this module, you will explore the following:

* Describing disease occurrence in populations
* Interpret epidemic curves and spatial data to describe disease occurrence in a population

## How does a veterinarian describe disease occurrence in populations?

Recognizing and considering disease occurs in populations of animals rather than in individual animals and describing disease occurrence in (animal) populations can be very useful in understanding the epidemiology of disease.

Typically, there are 3 ways to describe disease occurrence in populations - based on **time, space, and individual factors (from Module 1)**.

We will start by describing disease in animal populations across **time**.  The most common way to describe disease in time is by using **epidemiologic (epi) curves**.  A typical epi curve shows the occurrence of cases of disease (on the vertical axis) across time (on the horizontal axis). An excellent example can be found for the COVID-19 pandemic in the US on the [CDC COVID-19 Data Tracker](https://covid.cdc.gov/covid-data-tracker/#datatracker-home) website. Scroll down on the page to View Trends to visualize the epi curve for COVID-19 hospitalizations or deaths by week in the US.  In this format, it is easy to quickly understand how the disease is moving through the population.

Review the CDC Quick-Learn Lesson - [Using an Epi Curve to Determine Mode of Spread](https://www.cdc.gov/training/QuickLearns/epimode/)

This CDC website provides an excellent introduction to Epi Curves and shows why understanding the distribution of cases over time can provide inference about the pattern of transmission as well as the magnitude of the outbreak and likely time of exposure.

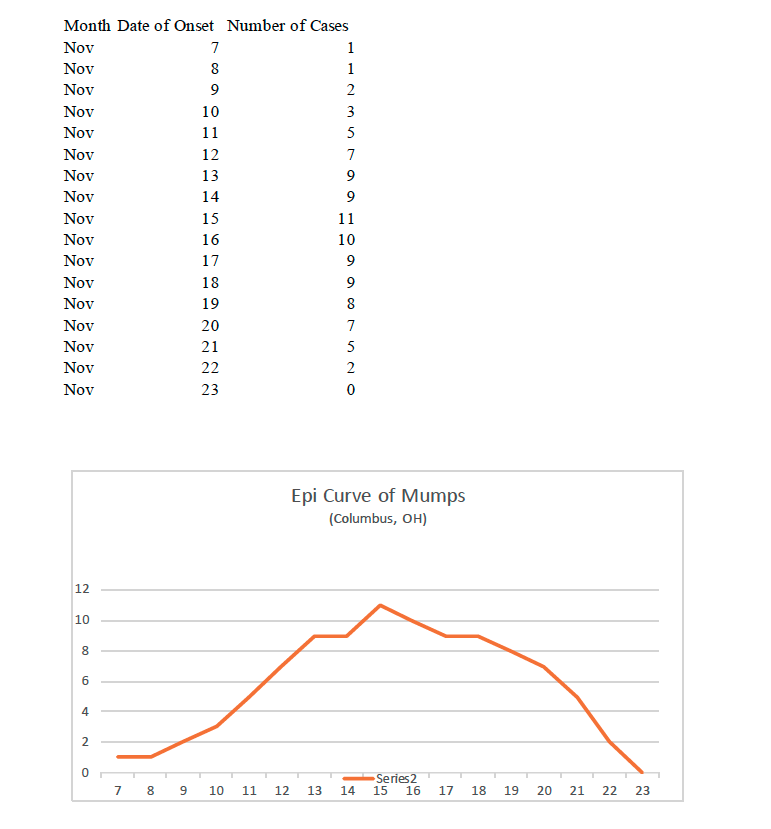
### How to Create an Epi Curve

Review the information in the How to Create an Epi Curve resource.

#### **Case 2.1. Use the Mumps in Columbus, OH** data,

located in information on how to create an Epi Curve, to create an Epi Curve.

Your Epi Curve should look like this:



## Interpreting epidemic curves and spatial data

In the previous module, you learned the importance of describing the level of disease occurrence in three ‘dimensions’:

* Time
* Place (spatial factors)
* Individual factors

All three of these factors are incorporated in the case definition, which is defined as a set of criteria used to decide whether an individual is considered as ‘diseased’ or ‘not diseased.’

### A case definition should include:

* Criteria related to time, place, and individual factors, plus
* Clinical features that make an individual a ‘case’, like clinical signs of disease and/or positive test results.

Once a proper case definition is set, we can detect if an epidemic is taking place. An **epidemic** is defined as **a sudden increase in the number of cases of a disease above what is expected for a given population and area**.

When information on disease occurrence, defined according to the ‘case definition’ and time when cases are declared/detected/ notified is available, we can build an epidemic curve as a visual display showing the frequency of new cases over time.

As we learned in the previous module, an epidemic curve is a way to show how the disease occurs through time.

Interpreting an epidemic curve also provides information about possible sources of the disease, based on classifying epidemics into either:

* **Common source epidemic**:

Epidemic in which all cases are exposed to a common source such as shared feed or water or air (may be continuous, intermittent or point source depending on the time that the exposure lasts).

* **Propagated epidemic**: Epidemic in which a case of disease serves as source of infection for subsequent cases, and the subsequent cases in turn serve as sources for later cases. This typically represents animal to animal transmission of infectious diseases.

### Case 2.2. Review the epi curve from another real-case outbreak in turkey operations:

A picture containing diagram, text, sketch

Description automatically generated

[Review the research paper found as a resource](https://vpmmanifold.com/projects/clinical-epidemiology/resource/wells-et-al)

In this outbreak of highly pathogenic avian influenza in Minnesota commercial poultry farms, you observe an initially infected flock during the week of March 1-8, followed by an increasing number of poultry flocks detected starting in late March with a peak number of detected cases during the week of April 19-26.  Later, smaller peaks of positive flocks occurred after that time until the week of May 31 to June 7.

This example shows multiple peaks of cases (of different sizes), which is characteristic of propagated epidemics seen in outbreaks of infectious diseases.  In this outbreak (part of a very large foreign animal disease outbreak in the US that year), the study investigators hypothesized that the outbreak may have been caused by avian influenza viruses from different sources and conducted a study that identified different factors associated with turkey farm-positivity based on the time of onset of disease.

Based on the pattern of presentation of the disease in a population (defined during a given time and space) using epi curves, we can differentiate between:

#### Endemic disease: Disease that occurs at a predictable frequency in a population.

#### Hyper-endemic disease: Disease constantly present at a high incidence and/or prevalence.

#### Sporadic disease: Disease that occurs occasionally in an infrequent and irregular manner.

#### Emerging disease: Disease that has appeared in a population for the first time or that is rapidly increasing in incidence or geographic range.

The question that now needs to be explored:  How does one interpret epidemic curves and classify epidemics into the above categories?

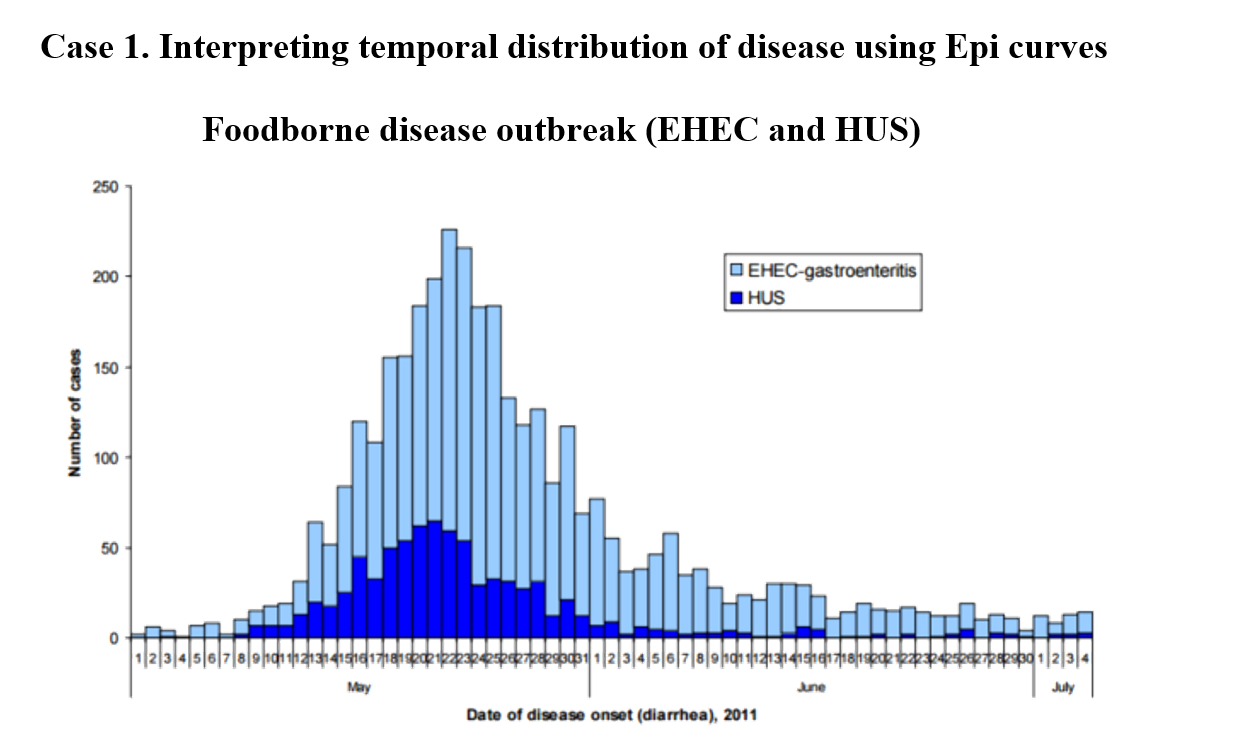
Assuming a careful description of cases to represent description of a single disease, the following questions are very helpful to describe the epidemiology of disease:

| **Question** | **Description of epidemiology** |
| --- | --- |
| What is the starting point of the number of cases? | If 0, then infers a new outbreak or newly detected disease.  If greater than 0, infers a baseline number of cases has been present which infers endemic disease. |
| How does the peak number of cases relate to the starting point or baseline? | If much higher peak, infers an epidemic (disease occurrence higher than normal or baseline) |
| How many peaks in the numbers of cases are detected? | A single peak infers a common source epidemic with individuals exposed at similar time to common source of infection.  Multiple peaks infers a propagated epidemic with multiple times of exposure to the source of infection. |
| If a propagated epidemic, what is the time frame between peaks of disease occurrence? | The time duration between the peaks provides an estimate of the incubation period of the disease (the time between exposure and onset of clinical disease). |

### Interpreting the temporal distribution of disease using Epi curves

Let's practice describing the epidemiology of disease using epi curves. Below are five scenarios -- first answer the questions based on your interpretation of the epi curve, then compare your answers to the correct answers.

#### Case 2.3. Foodborne disease outbreak - Enterohemorrhagic Escherichia coli (EHEC) and Hemolytic uremic syndrome (HUS)

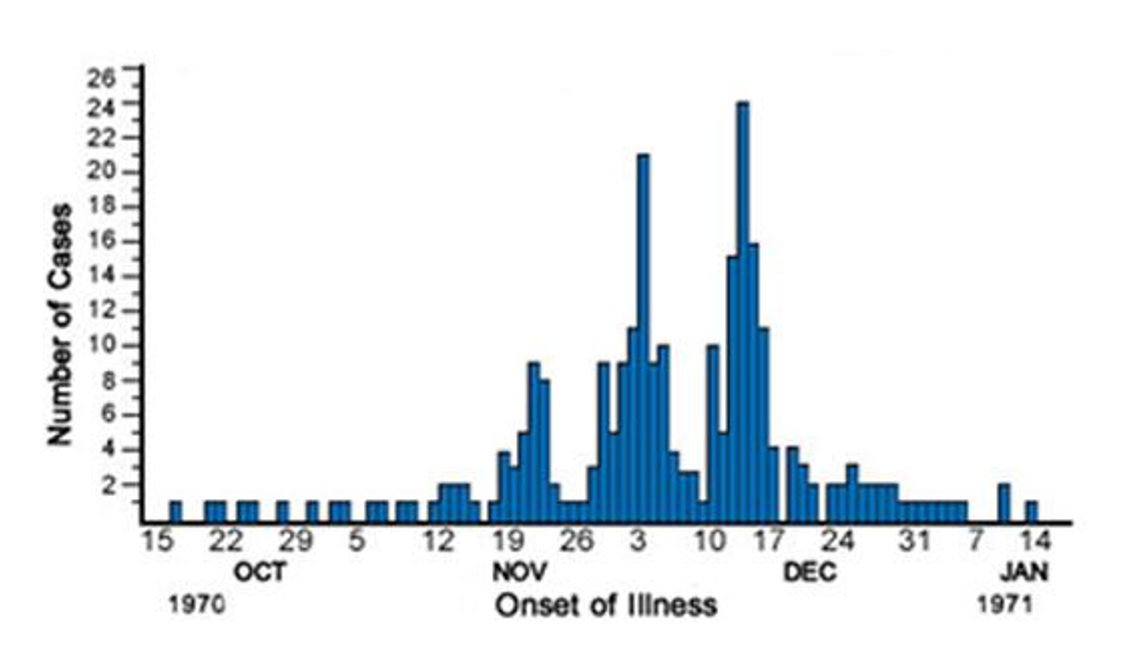


1. What is the starting point of number of cases?
2. How does the peak number of cases relate to the starting point or baseline?
3. How many peaks in the numbers of cases are detected?

**Answers:**

1. What is the starting point of number of cases?   
   0, which infers a new epidemic or newly detected disease.
2. How does the peak number of cases relate to the starting point or baseline?   
   Much higher peak infers an epidemic. A higher number of EHEC cases compared to HUS, though each shows a similar pattern of disease during a similar time frame (while a different peak number of cases), which infers a similar source of infection.
3. How many peaks in the numbers of cases are detected?   
   A single peak infers a common source (feed, water, or air source), with about 3 weeks from start (May 1) to peak of outbreak (May 21). The long tail of the epidemic (low numbers of cases detected) indicates some additional exposure to the source of infection through time.

#### Case 2.4. Another Disease Outbreak

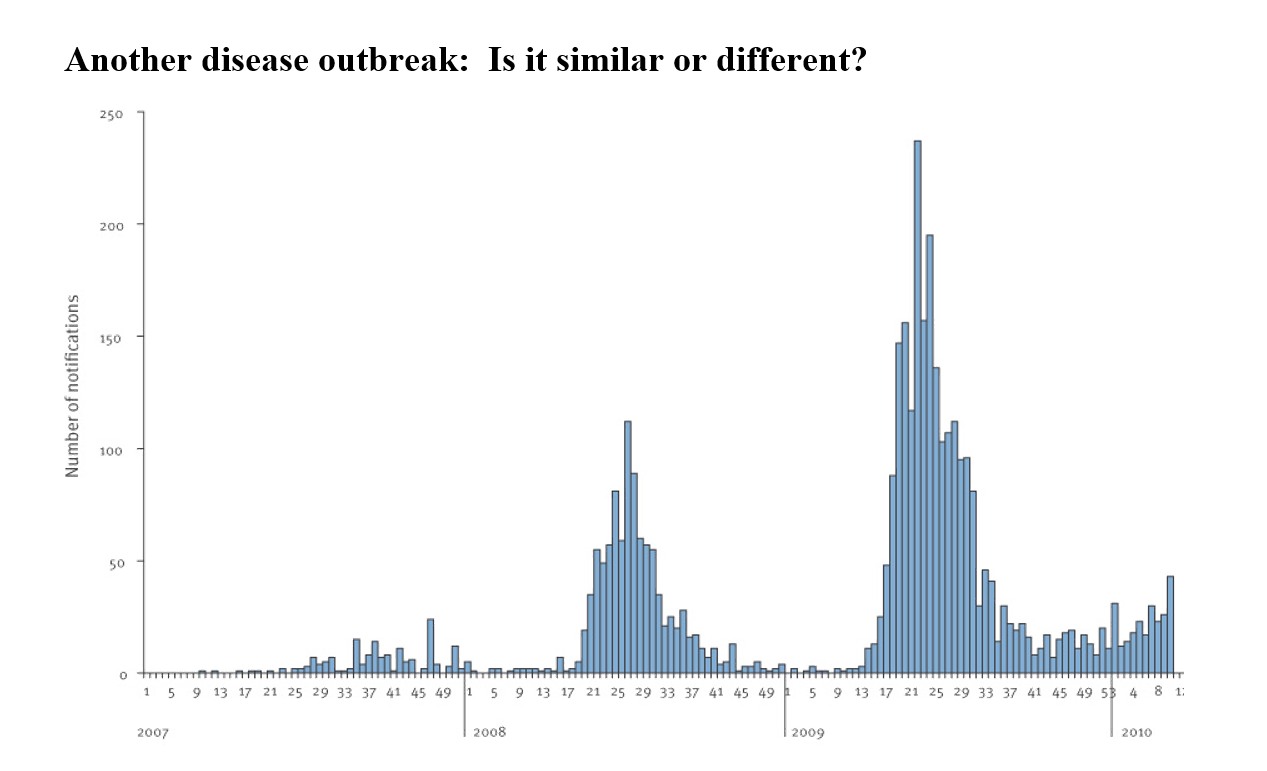


1. What is the starting point of the number of cases?
2. How does the peak number of cases relate to the starting point or baseline?
3. How many peaks in the numbers of cases are detected?
4. Since a propagated epidemic, what is the time frame between peaks of cases?

**Answers: Another Disease Outbreak**

1. What is the starting point of the number of cases?   
    O, which infers a new outbreak or newly detected disease.
2. How does the peak number of cases relate to the starting point or baseline?   
    A much higher peak compared to baseline infers an epidemic.
3. How many peaks in the numbers of cases are detected?   
    Multiple peaks infers a propagated epidemic (infectious disease).
4. Since a propagated epidemic, what is the time frame between peaks of cases?   
    About 2 weeks, which provides an estimated incubation period for disease.
5. Which diseases could this represent?    
    In humans, chicken pox and rubella (German measles) are infectious diseases with incubation periods similar to 14 days.

#### Case 2.5. Yet Another Disease Outbreak:  Is it similar or different?

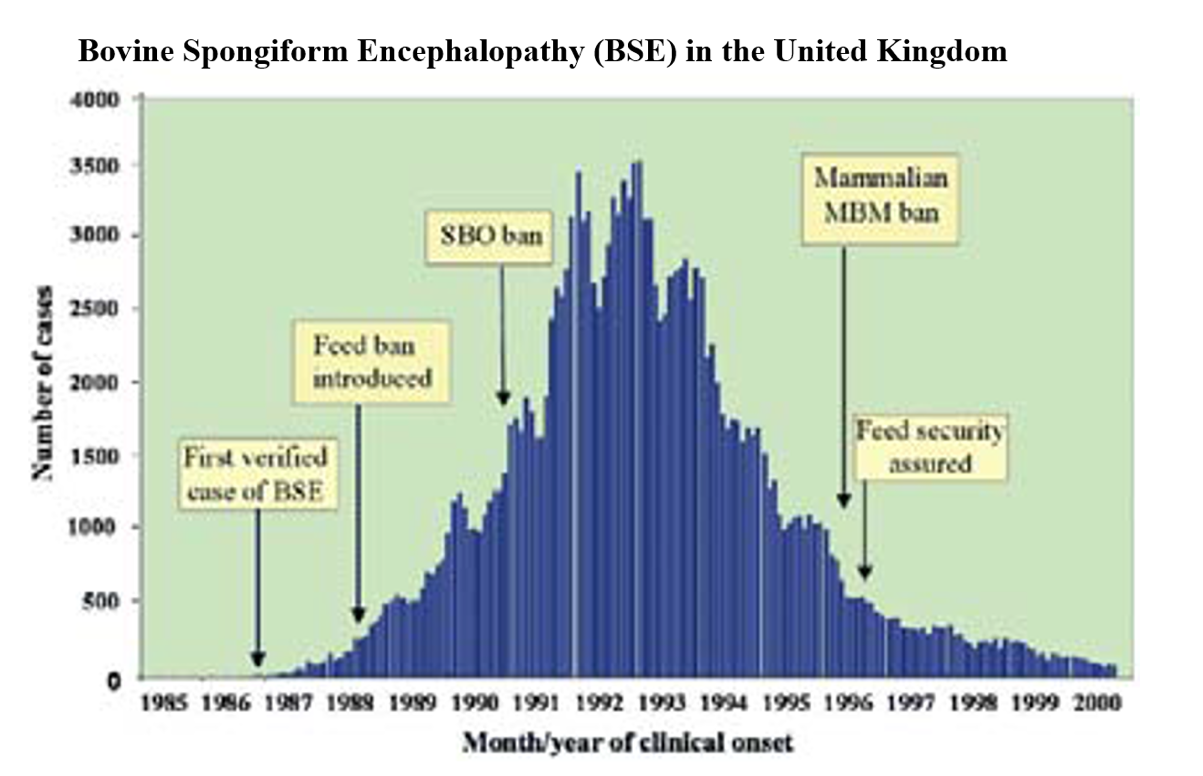


1. What is the starting point of number of cases?
2. How does the peak number of cases relate to the starting point or baseline?
3. How many peaks in the numbers of cases are detected?
4. Since a propagated epidemic, what is the time frame between peaks?

**Answers: Yet Another Disease Outbreak:  Is it similar or different?**

1. What is the starting point of number of cases?   
    0, which infers a new outbreak or newly detected disease.
2. How does the peak number of cases relate to the starting point or baseline?   
    Much higher peak compared to baseline infers an epidemic.
3. How many peaks in the numbers of cases are detected?   
    Multiple peaks infers a propagated epidemic (consistent with infectious disease).
4. Since a propagated epidemic, what is the time frame between peaks?   
    About 12 months. This type of disease occurrence can occur either due to a very long estimated incubation period for disease (one year) or a disease that occurs seasonally (summer weather).
5. Which human diseases occur seasonally in the summer?

#### Case 2.6. Bovine Spongiform Encephalopathy (BSE) in the United Kingdom

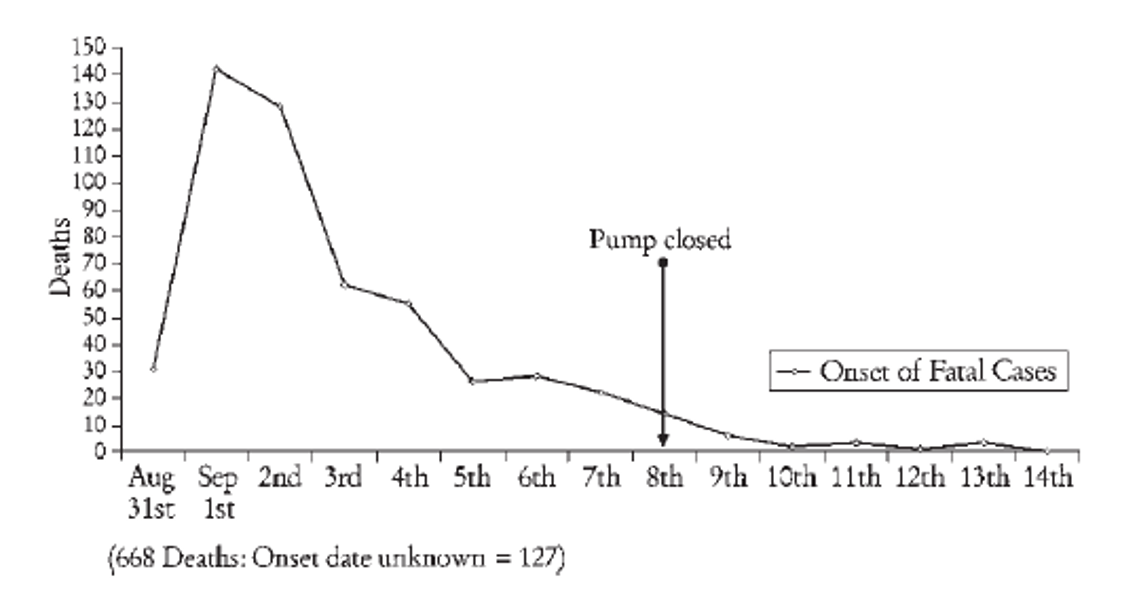


1. What is the starting point of number of cases?
2. How does the peak number of cases relate to the starting point or baseline?
3. How many peaks in numbers of cases are detected?

**Answers: Bovine Spongiform Encephalopathy (BSE) in the United Kingdom**

1. What is the starting point of number of cases?   
    0, which infers a new epidemic or newly detected disease.
2. How does the peak number of cases relate to the starting point or baseline?   
    A much higher peak infers an epidemic.
3. How many peaks in numbers of cases are detected?   
    A single peak infers a common source (feed, water, air source), with about 6-7 years from the start (1986) to the peak number of cases (1992-1993). The long tail of the epidemic (low numbers of cases detected) indicates some additional exposure to sources of infection through time.

#### Case 2.7. Cholera in London



1. What is the starting point of number of cases?
2. How does the peak number of cases relate to the starting point or baseline?
3. How many peaks in numbers of cases are detected?

**Answers: Cholera in London**

1. What is the starting point of number of cases?   
    30, which infers a baseline number of cases (endemic disease previously).
2. How does the peak number of cases relate to the starting point or baseline?   
    A much higher peak infers an epidemic detected.
3. How many peaks in numbers of cases are detected?   
    A single peak infers a common source of disease (feed, water, or air source), with about 3 days from the onset (August 31) to the peak number of cases (September 2). A long tail of the epidemic (low numbers of cases detected) indicates some additional exposure to sources of infection through time, which may include infectious disease transmission (propagated epidemic at later stages of outbreak).

### Interpreting the spatial distribution of disease

A method to describe the spatial distribution of disease is through the use of maps, highlighting the distribution of cases in space.  This can provide information on environmentally associated factors that may be related to disease which may suggest clues on possible sources, based on identification of local patterns and areas with higher risk.

#### Case 2.8. An example is the distribution of anthrax in Minnesota.

You can observe that cases have been identified in the western part of the state, up until 1999 in the southwest part of the state and more recently in the northwest part of Minnesota.  The distribution of these cases is consistent with the general area of the Great Plains in the state, an area formerly populated by grazing bison, and now grazed by cattle, sheep, and goats.  Due to the extreme environmental resistance of anthrax, spores are thought to survive for many decades in the environment, with cases occurring sporadically during certain environmental conditions.A map of minnesota with different colored squares

Description automatically generated with low confidence

1. Why do you think the cases have been observed in the northwest part of the state more recently?
2. What long term changes may explain this?

**Answers:**

1. Why do you think the cases have been observed in the northwest part of the state more recently?   
    It is not clear why anthrax cases in Minnesota have been reported only in the northwest part of the state since 2000, as bison and, more recently, cattle have been distributed all across the Great Plains in years past. Some have hypothesized that certain soil types and/or rainfall conditions may play a role.  Can you think of other ideas?
2. What long term changes may explain this?   
    Some have hypothesized that climate change may play a role in the longterm survival and/or detection of *Bacillus anthracis*, or perhaps in change of grazing patterns across the Great Plains.