# Chapter 2: Surveillance & Investigation

### Vocabulary List

* **Outbreak**: When there are more cases of a disease than expected in a certain area or population, usually in a short period of time.
* **Registries**: Lists or databases where information about certain diseases or conditions is kept to help researchers or health officials understand and track them.
* **Notification**: Informing health authorities about cases of a disease so they can take necessary actions.
* **Inclusion**: Adding something, like people or data, into a group or study.
* **Exclusion**: Leaving something out, like people or data, from a group or study.
* **Frequency**: How often something happens or how many times it occurs in a specific period.
* **Histogram**: A type of graph that shows the distribution of data by dividing it into intervals and displaying how many observations fall into each interval.
* **Incubation**: The period between when someone is exposed to a disease and when they start showing symptoms of it.
* **Baseline**: A starting point or reference level used for comparison in studies or research, often representing the initial state before changes occur.

## PART I

### Surveillance

(Bovbjerg, 2020)

The root word for epidemiology is epidemic. An epidemic is “an increase, often sudden, in the number of cases of disease above what is normally expected in that population in that area.” How do we know how much is expected? Surveillance! Public health surveillance is defined by the World Health Organization (WHO) as "The continuous, systematic collection, analysis, and interpretation of health related data needed for the planning, implementation, and evaluation of public health practice." Such surveillance can do the following:

1. Serve as an early warning system for impending public health emergencies
2. Document the impact of an intervention, or track progress towards specified goals
3. Monitor and clarify the epidemiology of health problems, to allow priorities to be set and to inform public health policy and strategies.

Surveillance activities can be either passive or active:

* In passive surveillance, the health department passively receives reports of suspected injury or illness. Think of this as waiting for disease reports to come to you. Many routine surveillance activities are passive. For instance, systems keeping track of communicable diseases, cancer, and injuries. Epidemiologists collect case reports that are sent to them by health care providers, laboratories, schools, or other entities that are required by law to report this information.
* In active surveillance, on the other hand, epidemiologists actively seek out cases of disease. For example, during an outbreak of salmonellosis associated with a specific source (for example, a restaurant), epidemiologists may contact health care providers in the area and ask each for a list of patients seen with symptoms consistent with salmonellosis. These patients are then contacted to see if they were exposed to the suspected source (the restaurant). National surveys are also considered active surveillance. The benefit of active surveillance is that it generally results in more complete data, while passive surveillance relies on others (who have numerous duties other than disease reporting) to report cases. The downside to active surveillance is that it is more resource-intensive, with increased personnel and financial requirements.
* Population Based: Some surveillance activities can be further characterized as population-based. The goal of population-based surveillance is to find every case that occurs within a population, and it is usually part of a mandated effort to collect cases of a specific condition of interest. An example of a condition for which we do population-based surveillance is cancer. Cancer registries aim to capture every case of cancer that occurs in the population the registry covers. This allows clinicians and public health professionals to monitor for trends in diagnoses that might signify a concerning change in the environment and/or for trends in survival that might follow improvements in treatment.

### Reportable Disease Notification and Reporting

In the United States, there is a list of conditions—mostly infectious diseases, but a few chronic diseases and injuries also make the list—that must be reported to the Centers for Disease Control and Prevention (CDC) whenever they are encountered by clinicians or health department officials. The World Health Organization (WHO) also has conditions that must be reported. For example, a patient visits a clinic complaining of high fever, cough, and watery eyes followed by a full-body rash. The clinician who sees the patient diagnoses measles. This clinic must then report the measles case to the local health jurisdiction, who in turn reports it to their parent agency, who in turn reports it to the CDC or WHO. This reporting ideally happens quickly, in a matter of days (or within hours for a potentially major threat).

This list of notifiable diseases is reviewed frequently and updated with current public health threats that are emerging (such as COVID-19 or Zika virus). Reporting is crucial to understanding trends of diseases and allowing epidemiologists to find cases of disease and investigate them to prevent future spread.

### Case Definition

(Dicker et al., n.d.-a)

Epidemiology relies on counting cases and making comparisons between groups. Before counting cases, the epidemiologist must decide what to count, that is, what to call a case. For that, the epidemiologist uses a case definition. A case definition is a set of standard criteria for classifying whether a person has a particular disease, syndrome, or other health condition. Some case definitions, particularly those used for national surveillance, have been developed and adopted as national standards that ensure comparability. Use of an agreed-upon standard case definition ensures that every case is equivalent, regardless of when or where it occurred, or who identified it.

Furthermore, the number of cases or rate of disease identified in one time or place can be compared with the number or rate from another time or place. For example, with a standard case definition, health officials could compare the number of cases of listeriosis that occurred in Lima, Peru in 2021 with the number that occurred there in 2019. Or they could compare the rate of listeriosis in Lima in 2021 with the national rate in that same year. When everyone uses the same standard case definition and a difference is observed, the difference is likely to be real rather than the result of variation in how cases are classified.

Other case definitions, particularly those used in local outbreak investigations, are often tailored to the local situation. For example, a case definition developed for an outbreak of viral illness might require laboratory confirmation where such laboratory services are available, but likely would not if such services were not readily available.

### Components of a Case Definition for an Outbreak

A case definition consists of clinical criteria and, sometimes, limitations on time, place, and person. The clinical criteria usually include confirmatory laboratory tests, if available, or combinations of symptoms (subjective complaints), signs (objective physical findings), and other findings. Case definitions used during outbreak investigations are more likely to specify limits on time, place, and/or person than those used for surveillance. Contrast the case definition used for surveillance of listeriosis (see box below) with the case definition used during an investigation of a listeriosis outbreak in Lima, Peru in 2000. 25,26

Both the national surveillance case definition and the outbreak case definition require a clinically compatible illness and laboratory confirmation of Listeria monocytogenes from a normally sterile site such as blood or cerebral spinal fluid (CSF), but the outbreak case definition adds restrictions on time and place, reflecting the scope of the outbreak.

### Listeriosis -- Surveillance Case Definition

**Clinical description**

Infection caused by Listeria monocytogenes, which may produce any of several clinical syndromes, including stillbirth, listeriosis of the newborn, meningitis, bacteriemia, or localized infections.

**Laboratory criteria for diagnosis**

Isolation of L. monocytogenes from a normally sterile site (e.g., blood or cerebrospinal fluid or, less commonly, joint, pleural, or pericardial fluid).

**Case classification**

Confirmed: a clinically compatible case that is laboratory confirmed.

Source: Centers for Disease Control and Prevention. Case definitions for infectious conditions under public health surveillance. MMWR Recommendations and Reports 1997:46(RR-10):49-50.

**Listeriosis -- Outbreak Investigation**

**Case definition**

Clinically compatible illness with L. monocytogenes isolated

* From a normally sterile site
* In a resident of Lima, Peru
* With onset between October 24, 2000 and January 4, 2001

A case definition may have several sets of criteria, depending on how certain the diagnosis is. For example, during an investigation of a possible case or outbreak of measles, a person with a fever and rash might be classified as having a suspected, probable, or confirmed case of measles, depending on what evidence of measles is present (see box below).

### **Measles (Rubeola) -- 1996 Case Definition**

**Clinical description**

An illness characterized by all the following:

* A generalized rash lasting greater than or equal to 3 days
* A temperature greater than or equal to 101.0 degrees fahrenheit (greater than or equal to 38.3 degrees celcius)
* Couch, coryza, or conjunctivitis

**Laboratory criteria for diagnosis**

* Positive serologic test for measles immunoglobulin M antibody, or
* Significant rise in measles antibody level by any standard serologic assay, or
* Isolation of measles virus from a clinical specimen

**Case classification**

Suspected: Any febrile illness accompanied by rash.

Probably: A case that meets the clinical case definition, has noncontributory or no serologic or virologic testing, and is not epidemiologically linked to a confirmed case. 

Confirmed

: A case that is a laboratory confirmed or that meets the clinical case definition and is epidemiologically linked to a confirmed case. (A laboratory-confirmed case does not need to meet the clinical case definition.)

Comment:

Confirmed cases should be reported to National Notifiable Diseases Surveillance System. An imported case has its source outside the country or state. Rash onset occurs within 18 days after entering the jurisdiction, and illness cannot be linked to local transmission. Imported cases should be classified as:

* International. A case that is imported from another country.
* Out-of-State. A case that is imported from another state in the United States. The possibility that a patient was exposed within his or her state of residence should be excluded; therefore, the patient either must have been out of state continuously for the entire period of possible exposure (at least 7-18 days before onset of rash) or have had one of the following types of exposure while out of state: a) face-to-face contact with a person who had either a probable or confirmed case or b) attendance in the same institution as a person who had a case of measles (e.g., in a school, classroom, or day care center).

An indigenous case is defined as a case of measles that is not imported. Cases that are linked to imported cases should be classified as indigenous if the exposure to the imported case occurred in the reporting state. Any case that cannot be proved to be imported should be classified as indigenous.

Source: Centers for disease control and Prevention. Case definitions for infectious conditions under public health surveillance. MMWR Recommendations and Reports 1997:46(RR-10):23-24.

A case might be classified as suspected or probable while waiting for the laboratory results to become available. Once the laboratory provides the report, the case can be reclassified as either confirmed or not a case, depending on the laboratory results. Case definitions can also change over time as more information is obtained, such as during the COVID-19 pandemic when we didn’t know much about the disease.

### Target Population and Sample

(Bovbjerg, 2020)

Epidemiologists concern themselves with populations, not individual people. The definition of a population is a group of people with a common characteristic. This could be residents of a country, people with type 1 diabetes, people under age 25 who work full-time, and so on. For epidemiologists, the population is the group of people about whom we wish to be able to say something. For instance, say that we are interested in whether the amount of sleep a student gets is related to their grade point average (GPA). If we are mainly interested in this relationship among college students, then our population might be full-time undergraduates. However, there are a lot of full-time undergraduates in the world; we cannot possibly enroll them all into our study. We therefore draw a sample from the target population and do the study with the people in the sample (which here will be some smaller group of full-time undergraduates).

Ideally, the sample will be similar enough to the target population that our results can indeed be generalized back to the target population; therefore, we would work to recruit a diverse sample of students who are similar to the population.  The generalizability of our sample does not always matter as much as it does in other fields.

### 

### Inclusion and Exclusion

We define populations via lists of inclusion and/or exclusion criteria. When defining a population, the list of inclusion and exclusion criteria must be sufficiently complete that any given person could look at it and decide whether they were in the population.

As an example, if we were planning a study of strength training to prevent osteoporotic fractures in elderly women, the inclusion and exclusion criteria would need to specify the following:

* The lower (and potentially, upper) age cutoff (in other words, what is elderly for our purposes?)
* Whether we are interested in biological females, those who identify as women, or both
* Whether there are any exclusions in terms of physical capabilities (for example, not all elderly women are able to do a strength training regimen)

We might choose to exclude women for whom exercise of any kind is contraindicated (for example, if they are heart failure patients), or those who have already had a hip fracture, and so on.

Note that when creating inclusion and exclusion criteria lists, only rarely is there a correct answer. Often, scientific or clinical considerations will help narrow it down, but in our example above, it probably doesn’t matter if we set the lower age bound at 60, 65, or 70, as long as we set one and stick to it.

## PART II

### Rates

(Dicker et al., n.d.-b)

In epidemiology, a rate is a measure of the frequency with which an event occurs in a defined population over a specified period of time. Because rates put disease frequency in the perspective of the size of the population, rates are particularly useful for comparing disease frequency in different locations, at different times, or among different groups of persons with potentially different size populations; a rate is a measure of risk.

 A rate may describe how quickly disease occurs in a population, for example, 70 new cases of breast cancer per 1,000 women per year. This measure conveys a sense of the speed with which disease occurs in a population, and seems to imply that this pattern has occurred and will continue to occur for the foreseeable future.

### Attack Rate

(Dicker et al., n.d.-c)

An attack rate is the proportion of the population that develops illness during an outbreak. For example, 20 of 130 persons developed diarrhea after attending a picnic.

### Cross Tabulation

(Bovbjerg, 2020)

Before getting into study designs and measures of association, it is important to understand the notation used in epidemiology to convey exposure and disease data: the 2 x 2 table. A 2 x 2 table (or two-by-two table) is a compact summary of data for 2 variables from a study, namely the exposure and the health outcome. Say we do a 10-person study on smoking and hypertension, and collect the following data, where Y indicates yes and N indicates no:

You can see that we have 4 smokers, 6 nonsmokers, 5 individuals with hypertension, and 5 without. In this example, smoking is the exposure and hypertension is the health outcome, so we say that the 4 smokers are exposed (E+), the 6 nonsmokers are unexposed (E−), the 5 people with hypertension are diseased (D+), and the 5 people without hypertension are nondiseased (D−). This information can be organized into a 2 x 2 table:

The 2 x 2 table summarizes the information from the longer table above so that you can quickly see that 3 individuals were both exposed and diseased (persons 1, 3, and 4); one individual was exposed but not diseased (person 2); two individuals were unexposed but diseased (persons 6 and 9); and the remaining 4 individuals were neither exposed nor diseased (persons 5, 7, 8, and 10). Though it does not really matter whether exposure or disease is placed on the left or across the top of a 2 x 2 table, the convention in epidemiology is to have exposure on the left and disease across the top.

When discussing 2 x 2 tables, epidemiologists use the following shorthand to refer to specific cells:

Or:

The margin totals are sometimes helpful when calculating various measures of association (and to check yourself against the original data).

### Histograms

(Dicker et al., n.d.-d)

A histogram is a graph of the frequency distribution of a continuous variable, based on class intervals. It uses adjoining columns to represent the number of observations for each class interval in the distribution. The area of each column is proportional to the number of observations in that interval. Figures 4.7a and 4.7b show two versions of a histogram of frequency distributions with equal class intervals. Since all class intervals are equal in this histogram, the height of each column is in proportion to the number of observations it depicts.

Figures 4.7a, 4.7b, and 4.7c are examples of a particular type of histogram that is commonly used in field epidemiology: the epidemic curve. An epidemic curve is a histogram that displays the number of cases of disease during an outbreak or epidemic by times of onset. The y-axis represents the number of cases; the axis represents date and/or time of onset of illness. Figure 4.7a is a perfectly acceptable epidemic curve, but some epidemiologists prefer drawing the histogram as stacks of squares, with each square representing one case (Figure 4.7b). Additional information may be added to the histogram. The rendition of the epidemic curve shown in Figure 4.7c shades the individual boxes in each time period to denote which cases have been confirmed with culture results. Other information such as gender or presence of a related risk factor could be portrayed in this fashion.

Conventionally, the numbers on the x-axis are centered between the tick marks of the appropriate interval. The interval of time should be appropriate for the disease in question, the duration of the outbreak, and the purpose of the graph. If the purpose is to show the temporal relationship between time of exposure and onset of disease, then a widely accepted rule of thumb is to use intervals approximately one-fourth (or between one-eighth and one-third) of the incubation period of the disease shown. The incubation period for salmonellosis is usually 12–36 hours, so the x-axis of this epidemic curve has 12-hour intervals.

The most common choice for the x-axis variable in field epidemiology is calendar time, as shown in Figures 4.7a, 4.7b, and 4.7c. However, age, cholesterol level, or another continuous-scale variable may be used on the x-axis of an epidemic curve.

In Figure 4.8, which shows a frequency distribution of adults with diagnosed diabetes in the United States, the x-axis displays a measure of BMI (body mass index). The choice of variable for the x-axis of an epidemic curve is clearly dependent on the point of the display. Figures 4.7a, 4.7b, or 4.7c are constructed to show the natural course of the epidemic over time; Figure 4.8 conveys the burden of the problem of overweight and obesity.

The component of most interest should always be put at the bottom because the upper component usually has a jagged baseline that may make comparison difficult. Consider the data on pneumoconiosis in Figure 4.9a. The graph clearly displays a gradual decline in deaths from all pneumoconiosis between 1972 and 1999. It appears that deaths from asbestosis (top subgroup in Figure 4.9a) went against the overall trend by increasing over the same period. However, Figure 4.9b (Bovbjerg, 2020) makes this point more clearly by placing asbestosis along the baseline.

Some histograms, particularly those that are drawn as stacks of squares, include a box that indicates how many cases are represented by each square. While a square usually represents one case in a relatively small outbreak, a square may also represent five or ten cases in a relatively large outbreak (Bovbjerg, 2020).

#### References

Bovbjerg, M. (2020, October 1). Foundations of Epidemiology. https://open.oregonstate.education/epidemiology/.

Dicker et al., Principles of Epidemiology in Public Health Practice, 3rd Edition.

Read this online at <https://books.byui.edu/pubh_370_readings/chapter_2_initiation_of_investigation>