Field Epidemiology
IN ACTION

Field Epidemiology Concepts

A Reference Guide for Field Epidemiologists

Field
Epidemiology
IN ACTION

Contents

Contents

Introduction

This booklet was produced by Field Epidemiology in Action for use by field epidemiology training program (FETP) fellows in Papua New Guinea and Solomon Islands. We also gladly share this with the broader field epidemiology community, and welcome people sharing this resource with their colleagues. Having shared language and understanding are important components in working effectively as a team; we hope that this booklet can be a useful tool to aid effective team communication.

How to use this boolet

This booklet does not include all terms used in field epidemiology; it is a selection of core epidemiology terms required for day-to-day work in the field.

Terms are organised by general category; it is important to note that many terms belong under several categories. In order to keep the document short and simple, we have defined each term only once. Where terms are mentioned in other sections, the word will be linked to the main definition (you can click on it to take you to the main definition).

Links to other sections of the document or to external websites are shown in blue text that is underlined.

This booklet has been designed for both intermediate and advanced FETP fellows, with more advanced content indicated by a green star ★ next to the term. Intermediate fellows are not expected to be familiar with the advanced content.

This will be a living document – updated periodically as required. The most recent version will be accessible at:

fieldepiinaction.com

Acknowledgements

The content of this booklet has been drawn from multiple sources. We would like to acknowledge and recommend the following epidemiology resources for reference or further learning. Please click on the title to be directed to the website.

United States Centers for Disease Control and Prevention, Principles of Epidemiology

United States Centers for Disease Control and Prevention guidance on evaluating surveillance systems

The Pacific Outbreak Manual

World Health Organization, Outbreak Toolkit

General Epidemiological Definitions

Epidemiology

The study of the **distribution** (who, where, when) and **determinants** (why, how) of **health related events** in human populations, and the application of this study to inform public health action.

Distribution

> Disease distribution refers to the frequency and pattern of health events in a population.

Frequency

Frequency refers to the number of health events (counts) in a population (e.g., the number of people living with HIV in a population at a particular time).

It also refers to the occurrence of a health event relative to the size of the population (rates). This measure allows comparison of disease frequency across different populations.

Pattern

Health events by time, place, and person. Time patterns may be yearly, monthly, weekly, daily, or hourly. They could also be seasonal, for example, vector-borne diseases patterns tend to be both seasonal and hourly, where risk is impacted by the season as well as the vector biting times (e.g., dusk). Place patterns can include differences between urban and rural geography, workplaces, altitude, or other geographic variations. Personal characteristics can include demographic factors (sex, age), socioeconomic status, behavioural factors or other environmental factors.

Determinants

> Refers to the causes or factors that bring about a change in a health event. Epidemiologists know that human health is influenced by a range of factors and that these factors can either increase or decrease the likelihood of experiencing a health event. These factors could be environmental (e.g., a weather event such as a tropical storm), social (e.g., socioeconomic status), behavioural (e.g., smoking or diet) as well as others.

Health event

In this booklet, ‘health event’ includes anything that could be considered to affect the health of a population. Often in **field epidemiology** this refers to communicable disease outbreaks, however, it can also refer to non-communicable disease, chronic disease, injury, occupational or environmental health, maternal and child health and any other health-related conditions.

Field epidemiology

The rapid, immediate investigation of urgent public health problems, such as **communicable disease** outbreaks. A primary goal of field epidemiology is to rapidly guide the selection and implementation of public health interventions to prevent negative health impacts, including death.

Case

A countable instance of the health condition under investigation (e.g., one person being diagnosed with a certain condition is a 'case'). These counts can be used for broade **public health surveillance** or during an outbreak response.

In the context of public health surveillance, a case is a single event meeting the identified case definition for a surveillance system. In a surveillance system, one person may contract a disease more than once and be counted as a case each time they contract that disease.

In an outbreak, a case is a single person meeting the outbreak case definition.

Case Definition

A set of standard criteria for deciding whether a person has a particular disease or health-related condition (to be classified as a case), by specifying clinical criteria and limitations on time, place, and person.

Surveillance system case definition

Include person, place, and time criteria, as well as clinical, laboratory (when available), and epidemiologic factors. Surveillance system case definitions remain consistent over time to enable quality comparison of data over time. Case definitions can change when there is a system change, such as the introduction of a new laboratory diagnostic test.

Example

> Case definition for influenza-like-illness for a **syndromic surveillance** system may be *'Fever ≥38°C AND cough AND onset within past 10 days'*

Example

> Case definition for COVID-19 for an **indicator-based surveillance** system at a hospital may be *'Laboratory confirmation of SARS-CoV-2'*

Outbreak case definition

Specific to each outbreak; they must specify person, place, time, and clinical criteria (signs and symptoms; laboratory diagnosis if available). Outbreak case definitions can change over time as more information is gathered throughout the outbreak investigation.

Example

> Following detection of a cluster of gastroenteritis cases from Village X on June 10, 2021, the initial case definition was: *Any person living in Village X presenting with diarrhoea from June 10, 2021*. Following further investigation of the initial cases, they had all attended a wedding party together. As a result, the case definition was changed to reflect this new intelligence: *Any person who attended the wedding party in Village X presenting with diarrhoea from June 10, 2021*.

Infectious (Communicable) Disease

Infectious disease

An infectious disease is an illness resulting from pathogens (e.g., bacteria, virus) that invade an organism (such as a human body), multiply, and cause infection. Communicable diseases are infectious diseases that can spread from one person or animal host to another. This is different to a non-communicable disease like diabetes which cannot be transmitted from one person to another.

Pathogen

A pathogen is an infectious microorganism that can cause disease or illness to its host, including virus, bacterium, protozoan, prion, and fungus.

Asymptomatic infection

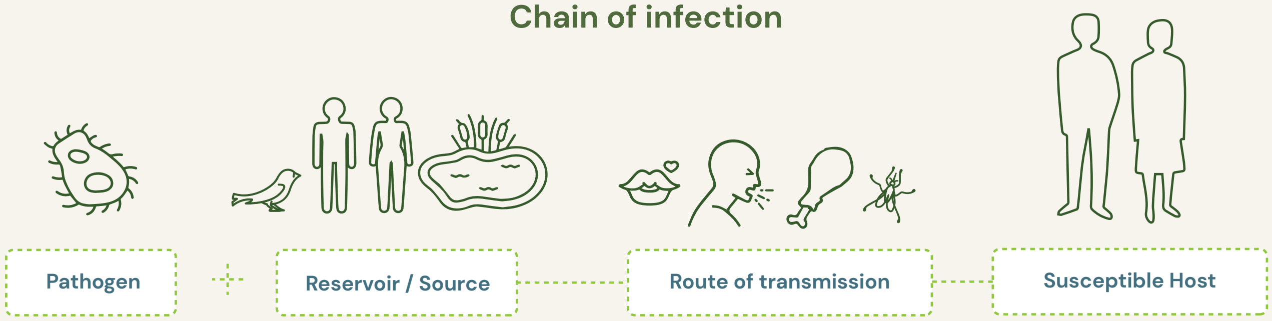
An infection without the presence of symptoms. Sometimes also called a 'subclinical infection'.

Chain of infection

A process that begins when an infectious agent leaves its reservoir or host through a portal of exit, and is conveyed by some route of transmission, then enters through an appropriate portal of entry to infect a susceptible host.

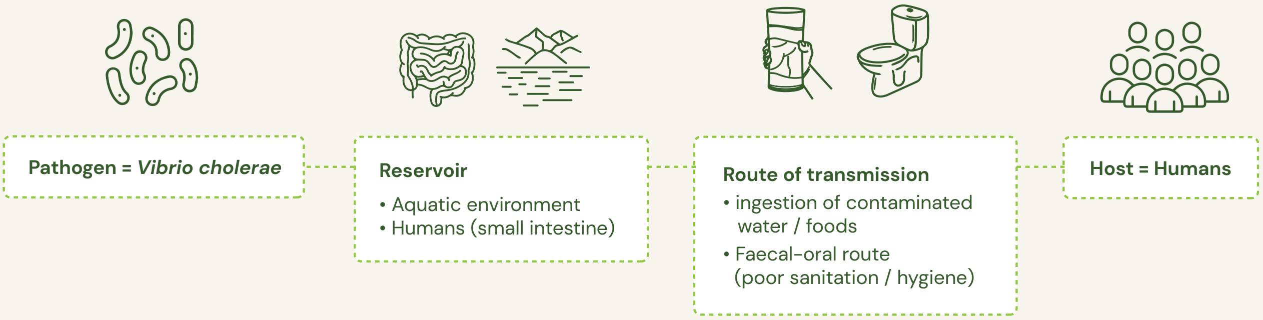
These principles are illustrated below:

Chain of infection



Pathogen + Reservoir / Source — Route of transmission — Susceptible Host

Example chain of infection for cholera



Pathogen = *Vibrio cholerae*

Reservoir
- Aquatic environment
- Humans (small intestine)

Route of transmission
- ingestion of contaminated water / foods
- Faecal-oral route (poor sanitation / hygiene)

Host = Humans

Route of Transmission

An infectious agent may be transmitted from its natural reservoir to a susceptible host in different ways. There are different classifications for modes of transmission:

Direct transmission

The immediate transfer of an agent from a reservoir to a susceptible host by direct contact or droplet spread.

Direct contact

> Occurs through skin-to-skin contact, kissing, and sexual intercourse. Direct contact also refers to contact with soil or vegetation that has infectious organisms.
>
> - **Example:** Gonorrhoea is spread from person to person by direct contact.
> - **Example**: Hookworm is spread by direct contact with contaminated soil.

Droplet transmission

> Refers to spray with relatively large, short-range aerosols produced by sneezing, coughing, or even talking. Droplet spread is classified as direct because transmission is by direct spray over a short distance (1–2 metres), before the droplets fall to the ground.
>
> - **Example**: Pertussis and meningococcal infection are examples of diseases transmitted from an infectious patient to a susceptible host by droplet spread.

Indirect transmission

The transfer of an infectious agent from a reservoir to a host by suspended air particles, inanimate objects (vehicles), or animate intermediaries (vectors).

Airborne transmission

> Occurs when infectious agents are carried by tiny, lightweight dust or droplet nuclei suspended in air. Airborne dust includes material that has settled on surfaces and become resuspended by air currents/breeze, as well as infectious particles blown from the soil by the wind. Whilst droplets are larger and heavier and fall to the ground quickly, with airborne transmission, the agent may remain suspended in the air for long periods and can be blown over long distances.
>
> - **Example:** Measles is an airborne pathogen.

Vehicle borne

> Vehicle transmission is the indirect transmission of pathogens through vehicles such as water, food, biological products such as blood, and fomites (inanimate objects such as handkerchiefs, bedding, or surgical scalpels).
>
> - **Example:** A vehicle may passively carry a pathogen — such as food or water may carry *Vibrio cholerae*. Alternatively, the vehicle may provide an environment in which the agent grows, multiplies, or produces toxin, such as improperly canned foods being an environment that supports production of botulinum toxin by *Clostridium botulinum*.

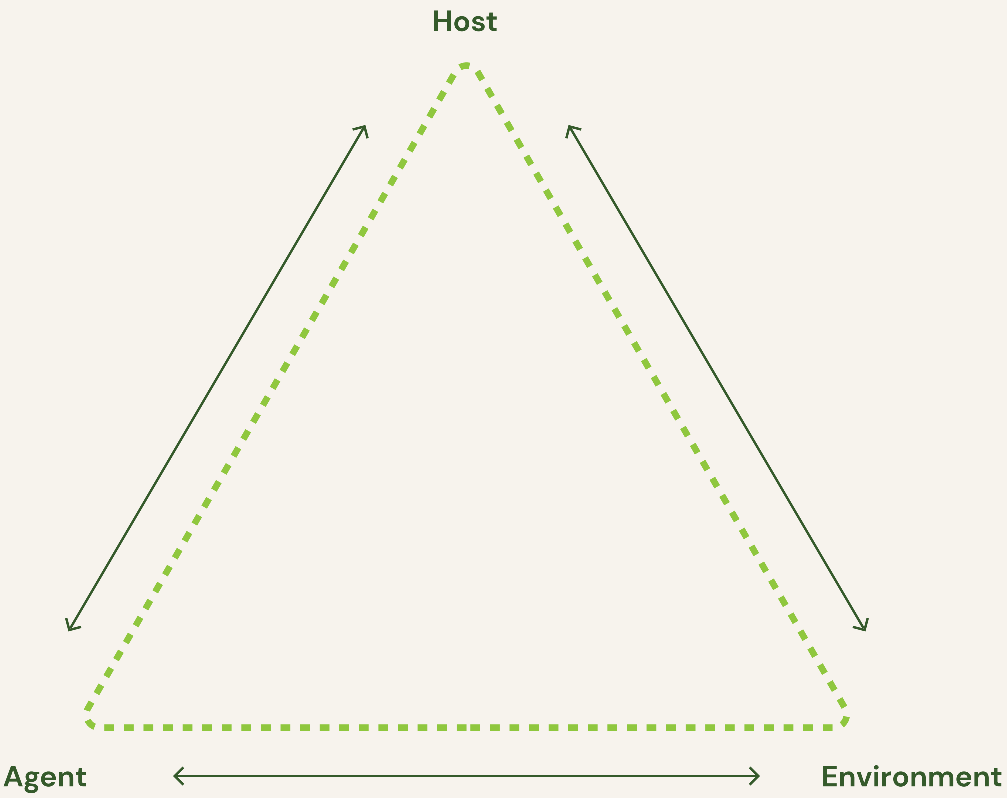
Vector borne

> Vectors such as mosquitoes, fleas, and ticks may carry an infectious agent through purely mechanical means or may support growth or changes in the agent.
>
> - **Example:** Flies can carry *Shigella* on their feet and fleas carry *Yersinia pestis*, the causative agent of plague, in their gut. In contrast, in biologic transmission, the parasite that causes malaria, for example, matures in an intermediate host (the mosquito) before it can be transmitted to humans.

Epidemiologic Triangle

The traditional model of infectious disease causation includes three components: an external agent, a susceptible host, and an environment that brings the host and agent together, so that disease occurs.

When we think about diseases or **health events** occurring, we always need to think about the interactions between the host, the environment, and the agent. This is called the epidemiologic triangle. This epidemiologic triangle is important in helping us think about how diseases are being spread and where we can intervene to stop transmission. Sometimes there are multiple interventions that we can take, and we need to consider the benefits of each one.



Agent (pathogen):

The agent is what causes the disease. In terms of infectious disease, the agent is typically a microbe, such as bacteria, virus, fungi, or protozoa. Commonly, people refer to these as “germs”.

- **In epidemiology, we want to know:**
  What do we know about this agent?
  How does it spread?
  How long does it survive in the environment?

Susceptible host:

Hosts are usually humans or animals who are exposed to and carry a disease. Sometimes the host gets sick, or sometimes the host can carry the agent to infect others, but not get sick themselves. Susceptibility of the host to the agent is dependent upon many factors, including genetics, previous immunity, pregnancy status, nutritional status, and comorbidities. Two doses of the measles, mumps, and rubella (MMR) vaccine, for example, is approximately 97% effective at preventing measles infection. Someone who has received these two doses, for example, is less susceptible to measles infection than someone with one or no previous MMR vaccines.

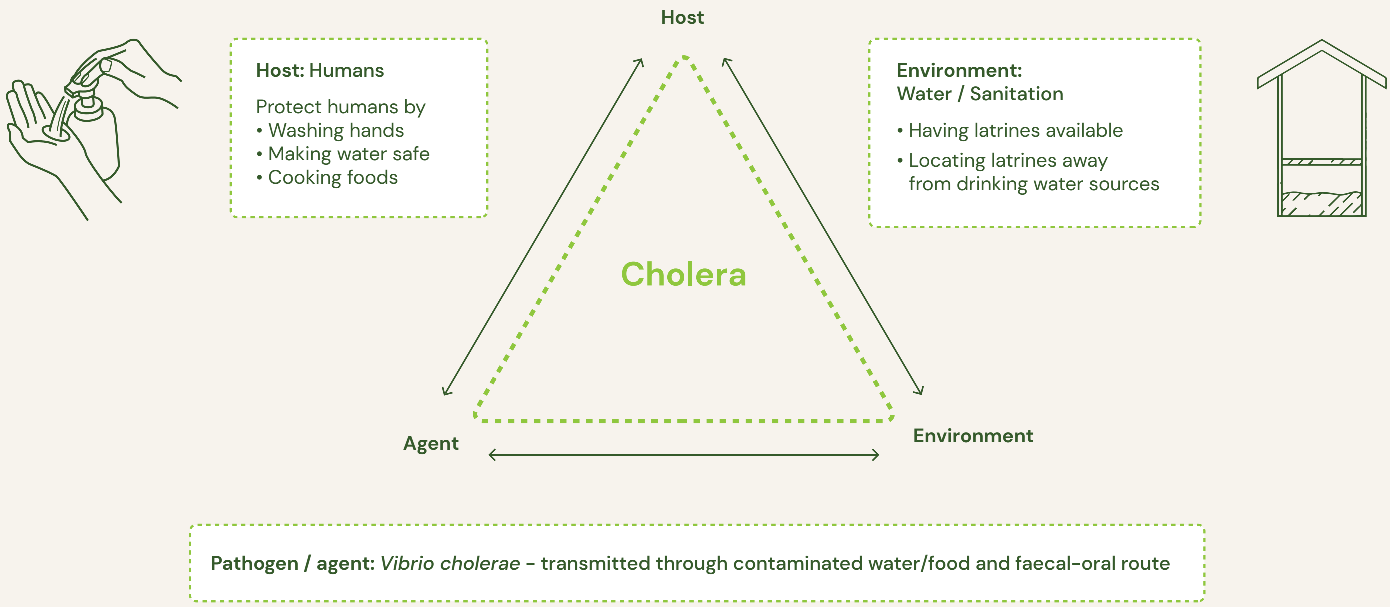
- **In epidemiology, we want to know:**
  What is different about the people who are affected?
  Are they elderly or immunocompromised?
  Do they have another health condition?
  Are they vaccinated?
  Why are they more likely to get infected?

Environment:

The environment is the surroundings, external to the host, that cause or facilitate transmission of the disease. For influenza, the environment conducive to transmission is the cooler winter weather. For the cholera bacterium, the environment that facilitates transmission is water sources contaminated with infected faeces.

- **In epidemiology, we want to know:**
  What is different about the environment where transmission is occurring?
  Are we in a city where people are in close contact?
  Are we in a rural area where people have more contact with animals?
  Are we in a coastal region where people have contact with water?
  Is the water source potentially contaminated?

Example of the epidemiologic triangle and how we can use this to consider control measures for cholera:



Host: Humans

Protect humans by
- Washing hands
- Making water safe
- Cooking foods

Environment:
Water / Sanitation

- Having latrines available
- Locating latrines away from drinking water sources

Pathogen / agent: *Vibrio cholerae* – transmitted through contaminated water/food and faecal-oral route

Stages of Disease

By knowing the stages of disease, we can predict when people are infectious (which can help us put in place control measures) and who else is likely to have been exposed (e.g., family members).

Incubation period

The incubation period is the time between being exposed to a pathogen and the onset of symptoms of disease. During this time, pathological changes are occurring in the host but they are unaware of them; it is sub-clinical. The incubation period is different for each disease and is usually given as a range.

- **Example:** The incubation period of COVID-19 is typically 2–14 days. The mean (average) incubation period, however, is 5 days, with some variants having shorter incubation periods.

Latent period

The time between exposure and infection, also known as the pre-infectious period. The pathogen is present, but there are no signs and symptoms in the host.

Clinical disease

The time of the disease process where there are signs and symptoms. It begins with the onset of symptoms and ends with recovery, disability, or death.

Infectious period

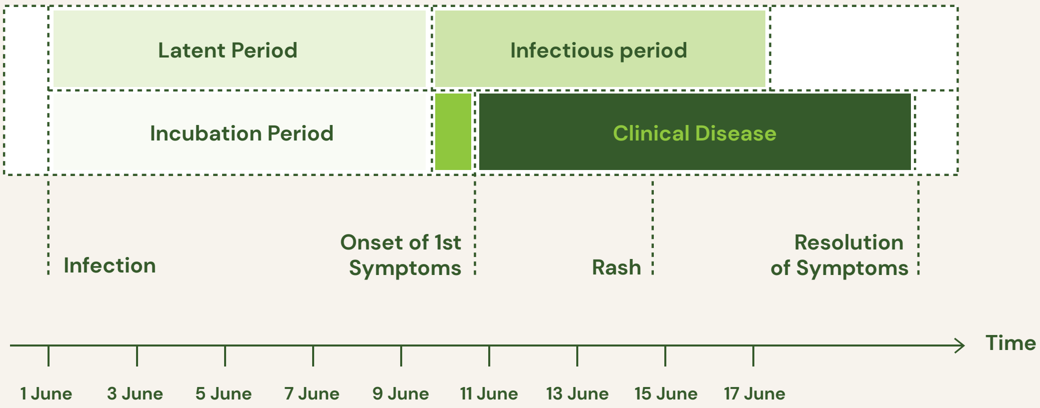
The infectious period is the time from when an infected person is able to pass the agent onto another person, until they are no longer able to pass the infectious agent on to another. The host may become infectious at any point after infection. This time varies with each pathogen.

Often, a person is infectious once they have symptoms, such as with pertussis/whooping cough. Sometimes, however, people can be infectious without symptoms or before symptom onset, such as with COVID-19 where a person is infectious up to 48 hours before symptom onset.

Rapid contact tracing is particularly important in outbreaks of diseases where people are infectious prior to symptom onset, as they are likely to have been living their life normally at work, school, socially, and home – and therefore exposing many people – before realising that they are infectious.

- **Example:** In the case of measles, the person is infectious up to four days before a rash appears and, in this example below (as indicated in the **BRIGHT GREEN** part of the image), one day before onset of any symptoms.

Example of the stages of disease below, using measles as an example:



Public Health Surveillance

Surveillance is health related data for action.

It is the systematic ongoing collection, collation, and analysis of data and the timely dissemination of information to those who need to know so that action can be taken.

Systematic

Surveillance is organised in a process and has rules about how it operates

Ongoing

Happening all the time. There is no end-date to surveillance.

Public health surveillance data can be collected for many different health events. These include communicable and non-communicable diseases, food or drug safety, environmental factors, and more.

There are different types of public health surveillance (definitions provided below). Often, one surveillance system will have multiple types working together at the same time.

Surveillance cycle

There are four key components of the surveillance cycle (pictured below)

Data collection

Relevant health data are collected about people who meet a case definition

Data management and analysis

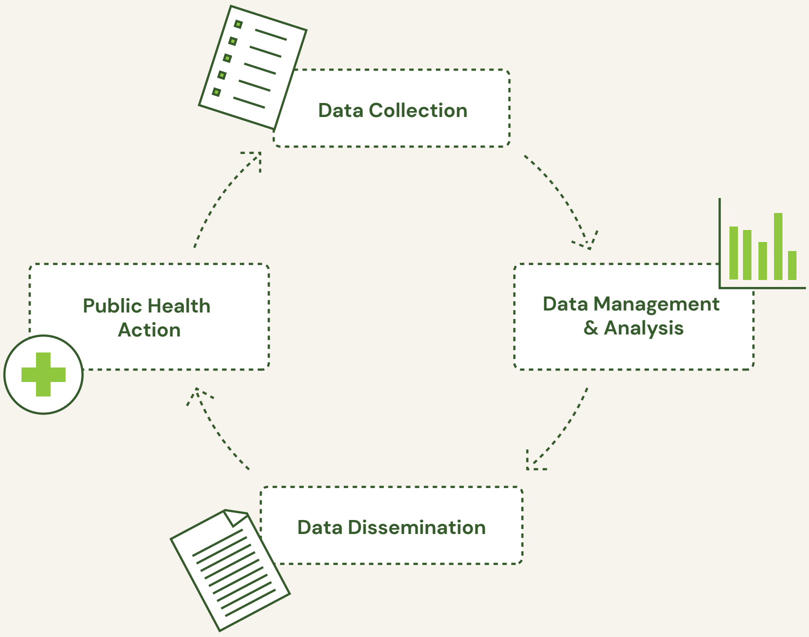
The health data are collated (combined and organised) and regularly analysed to see if there is something different compared with normal

Data dissemination

Regularly reporting the findings of data analyses back to the data collectors, the community, and those responsible for policy or public health action

Public health action

The data are used as an evidence base to take action to control or prevent the spread of disease



Surveillance system objective

Every surveillance system must have clear objectives (reasons for collecting data/goals to achieve). These objectives should accompany a clear description of how data are collected, collated, and analysed, and how the data will be used to prevent or control disease. A system will often have more than one objective. The surveillance system objectives will determine the most critical system attributes to incorporate into system design, monitoring, and evaluation.

Example: An objective of a tuberculosis surveillance system may be to identify persons with active disease to ensure that they are effectively and efficiently treated.

Example: An objective of a sexually transmitted infection surveillance system might be to inform testing and communication strategies which aim to identify and prevent disease.

Active surveillance

Active surveillance collects data by actively going out to find new cases. This means the health department takes the initiative to contact health providers, call or visit health facilities, and/or review records. It can also involve health workers going into community to systematically try to find cases by going house to house.

Active surveillance is often used when a disease is targeted for elimination (e.g., polio), or during the early stages of outbreak investigations (e.g., COVID-19 response).

Example: The central level may call provincial hospitals every day to find out if there have been any new COVID-19 cases present.

Passive surveillance

Passive surveillance involves the regular reporting of disease data by health care providers based on a known process. When patients who meet a case definition present to health facilities, passive surveillance relies on this information to be entered into the system. The reporting site then sends the information to the higher health system level. There is no active search for cases.

Example: Reviewing the clinical register at a participating health facility for the number of people who meet the surveillance case definition for diarrhoea, then filling out the weekly surveillance report form, and finally reporting the numbers to the national level.

Indicator-based surveillance

Indicator-based surveillance is routine, structured reporting of cases of disease or syndrome, typically from a health facility. Reporting is regular and ongoing, following an agreed process, often on a weekly or monthly basis.

Example: Information obtained through indicator-based surveillance might be reports received on a regular basis and entered routinely into a weekly surveillance report form or a database on the number of laboratory-confirmed cases of malaria identified at a hospital laboratory.

Syndromic surveillance

Syndromic surveillance uses **case definitions** of clinical signs and symptoms (syndromes) rather than a laboratory diagnosis. This means that identifying cases can be rapid as it is based on clinical signs and symptoms that can be done quickly at a clinic, rather than needing to wait for a laboratory confirmation.

It is useful for early detection of outbreaks and clusters where there may not be laboratory capacity to diagnose cases. This then enables a rapid response to initiate control measures.

Example: Using a syndromic case definition of '*Acute fever ≥38°C and non-vesicular rash*' enables healthcare workers to quickly identify potential measles cases if there is not laboratory capacity to test. While a laboratory test is subsequently recommended to confirm the diagnosis, by using the syndromic case definition, control measures can be rapidly initiated as though it is a true measles case to prevent onward transmission.

Sentinel surveillance

Sentinel surveillance is a sample of surveillance sites that contribute reports to the system. Sites may be chosen due to geographical spread, capacity to diagnose or test, or for financial reasons (i.e., Not able to include all sites). Due to being a sample of sites, the data are not representative of the entire population, but are an indicator of what is happening with particular diseases or syndromes in those select areas.

Example: Having people with microscopic examination skills at every facility in a country may not be possible. Instead, hospitals with suitably trained laboratory staff may be a sentinel site for malaria surveillance. This sentinel surveillance system will not be able to provide fully representative data of the malaria situation for the whole country, but it can identify trends to the particular catchment areas that those hospitals serve.

Event-based surveillance

Event-based surveillance is the rapid, unstructured capture of unusual events that may impact public health, such as unusual disease patterns, animal disease or die-offs, chemical spills or environmental contamination. These reports may come from formal channels, such as from healthcare workers at facilities, or from informal channels, such as local media and community members reporting to facilities or to dedicated hotlines. Events detected by event-based surveillance require urgent verification and response.

Example: A farmer calls a dedicated hotline to report that 70 of their 80 chickens died suddenly and unexpectedly this morning.

Example: A doctor calls the health department directly to alert them that they have seen six community members that morning from the same village with signs of neurotoxicity, and normally they would have none.

Surveillance system attributes

The attributes of a surveillance system affect the ability of a system to meet its objectives and the usefulness of the resulting information for public health action. Depending on the objectives of a surveillance system, some attributes will be more important than others.

Surveillance systems can be described by some of the following **attributes:** acceptability, data quality, flexibility, simplicity, stability, representativeness, and timeliness. The below descriptions of the attributes are sourced from the **Centers for Disease Control and Prevention guidance on evaluating surveillance systems**.

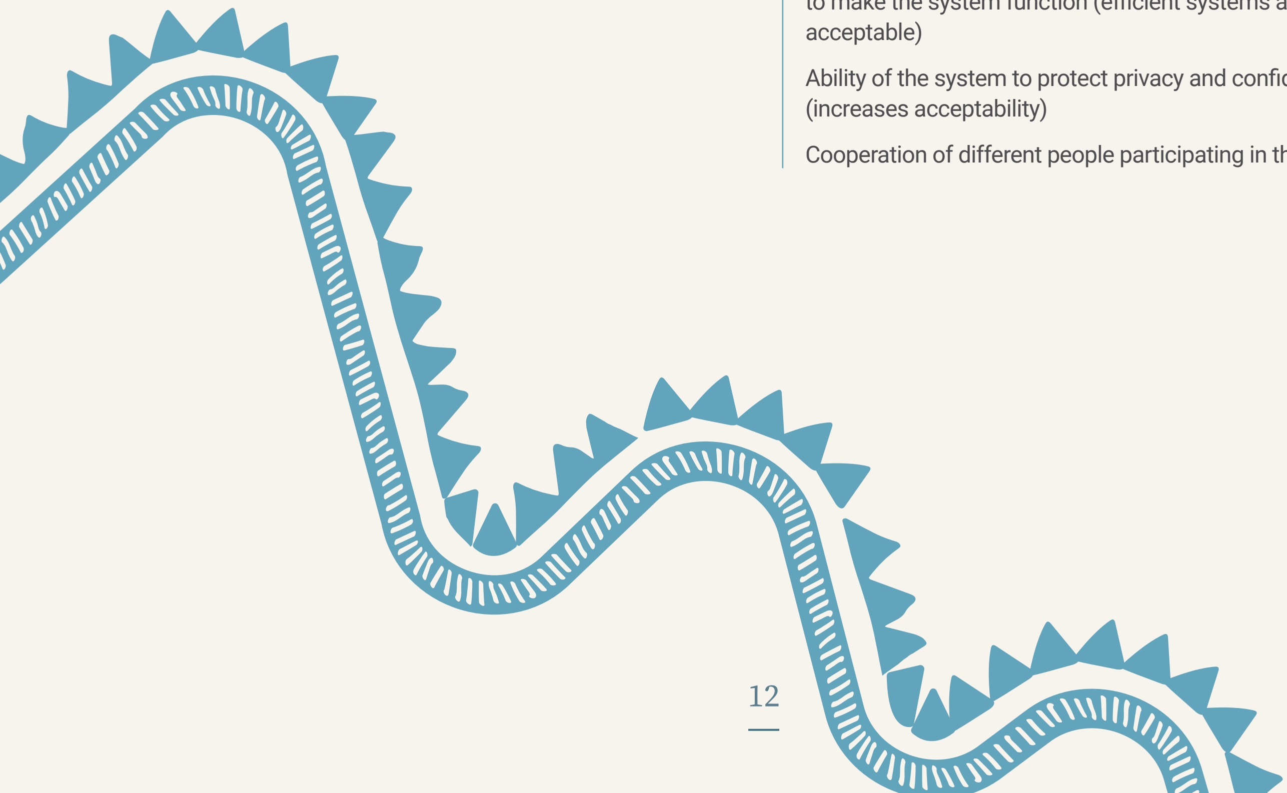
A surveillance system's objectives frame the way in which we **monitor and evaluate** a system's performance over time. Where timeliness is a particularly critical attribute for an event-based surveillance system, for example, measures of time between each step in the system would be an important factor to monitor on an ongoing basis, and to evaluate periodically.

Acceptability

Willingness of persons and organisations to participate in the surveillance system. It is related to whether users perceive that the information collected serves the surveillance system's goal.

Some factors influencing the acceptability of a particular system include:

> The public health importance of the event (if an issue is considered to be important to public health, people may be more accepting of the system)
>
> Acknowledgement of the data collector's contribution (people want to feel that their effort is meaningful and valuable)
>
> Dissemination of data back to reporting sources and stakeholders (contributors and next-users can understand how the data are used for decision-making
>
> How much time, effort, and difficulty is required to make the system function (efficient systems are more acceptable)
>
> Ability of the system to protect privacy and confidentiality (increases acceptability)
>
> Cooperation of different people participating in the system



Data quality

Data quality is measured by the completeness and validity of data in the surveillance system.

Completeness is related to whether there are missing records or data fields. It is often measured by the number or proportion of unknown or blank responses to variables in the system. The term "*zero reporting*" refers to reporting when there are no known cases. Data completeness is important to recognise a true zero case, rather than a "blank" response.

Validity is the capacity of the system to capture the true value and data specifications (e.g., consistency of reporting M/F instead of male/female). It is often related to how many errors can be found in the data in the system.

Flexibility

A flexible public health surveillance system can adapt to changing information needs or operating conditions with little additional time, personnel, or allocated funds. Flexible systems can accommodate new health-related events, changes in case definitions or technology, and variations in funding or reporting sources.

Example: Existing respiratory surveillance systems (e.g., For influenza-like illness or severe acute respiratory infections) being adapted for COVID-19, shows systems that were flexible to being adapted in a short time.

Simplicity

Structure and ease of operation of the system across the surveillance cycle from data collection to public health action. Systems should be as simple as possible while still meeting their objectives.

Simplicity may be related to:

The amount and type of information needed to decide if case definition has been met (e.g., may be a simple syndromic case definition or rapid test, or more complex clinical and laboratory diagnosis)

The amount and type of data needed for cases (e.g., demographic data are simpler than needing many items of risk factor data, such as occupation and travel history)

The method of data collection, including the time spent on collecting data or training staff in collecting data (e.g., collecting data from one source is simpler than needing to collate data from multiple sources)

The method of managing data (e.g., whether paper-based or electronic databases are easier to navigate)

The methods for analysing and interpreting data and preparing reports for dissemination (e.g., It is simpler if one analysis can be used for all reporting, rather than separate analyses)

Stability

Ability to collect, manage, and provide data without failure (reliability) and to be operational when needed (availability). A lack of dedicated resources might affect the stability of a public health surveillance system.

An unreliable system could be a system where there are insufficient staff to consistently be able to detect cases or analyse data.

Representativeness

Ability of the system to accurately describe the occurrence of a health condition under surveillance over time and its distribution in the population by place and person.

To generalise findings from surveillance data to the population at large, the data from a public health surveillance system should accurately reflect the characteristics of the health-related event under surveillance. These characteristics generally relate to time, place, and person.

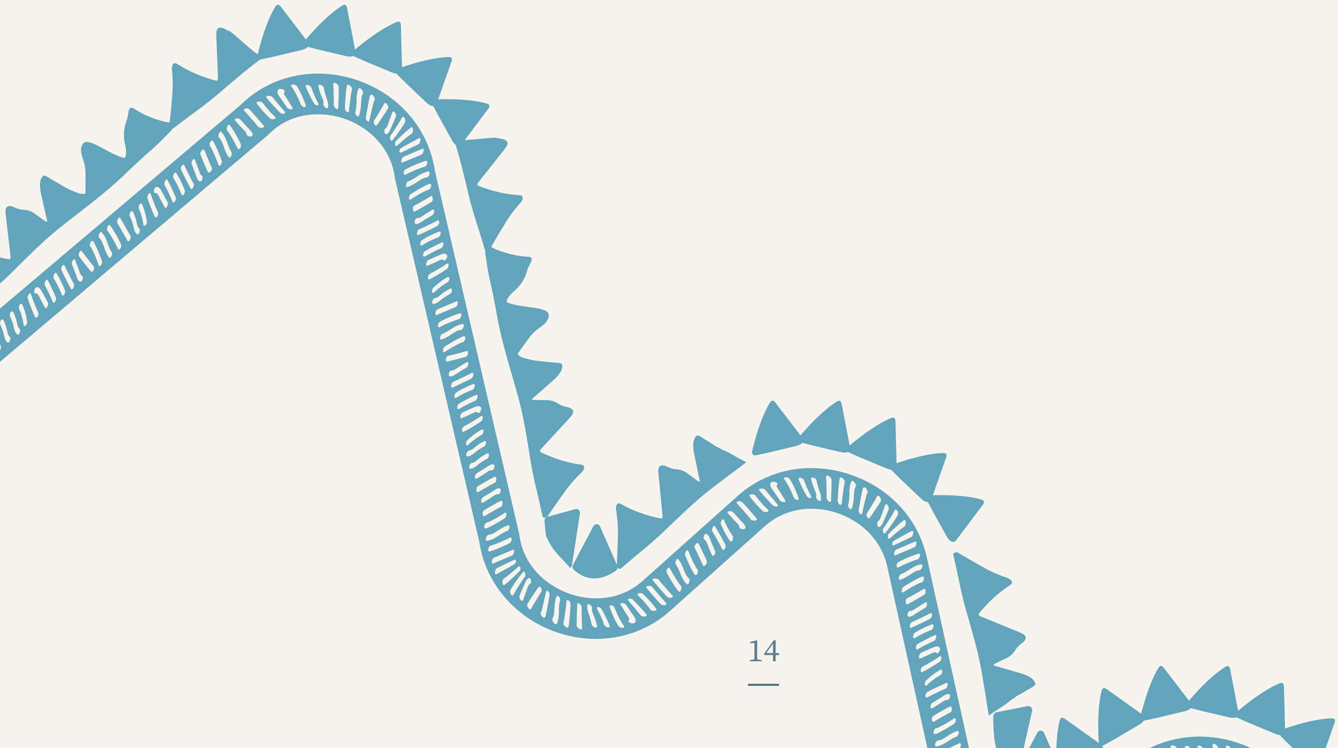
Surveillance data from all health facilities across the country may be more representative than data from two sentinel sites, for example.

Timeliness

Time between any two steps in the surveillance system. The importance of timeliness between each step in the surveillance system depends on the system's objectives and health event under surveillance.

> **Example:** for a surveillance system with a primary objective of detecting outbreaks, timeliness is crucial as it is important to implement control efforts quickly. There needs to be a short time between detecting a case or an outbreak and public health response.

> **Example:** for a surveillance system with a primary objective of monitoring incidence of disease over time for an annual report, timeliness is less important and there is likely to be more time between steps in the surveillance cycle.



Outbreak Response

Outbreak

A higher number of cases of a disease above what is normally expected in a certain population in a certain area. Some diseases are so rare and have such large public health consequence that a single case can be classified as an outbreak (e.g., Polio). Usually these changes are sudden, however, some disease outbreaks occur slowly due to the transmission of the disease (e.g., Syphilis).

Outbreak steps

The following table shows the eight steps in an outbreak investigation with an overview of each step. You do not have to complete the steps in order.

Outbreak Step	Overview of step
1. Confirm the outbreak	Questions to consider to verify an outbreak: • Is the number of cases more than expected? (e.g., Compared with previous weeks' data? Compared with same period in previous year? Considering seasonality?) • Check accuracy of initial reports • Do cases have the same disease? • Has there been a change in testing and/or reporting of the disease? (e.g., Increased local awareness leading to increased testing or reporting? Increase in population size? New test introduced? New surveillance site(s) added? Change in case definition?)
2. Establish a diagnosis (if you can)	• A specific diagnosis is useful but must not delay the outbreak investigation • Obtain clinical details from health workers • Laboratory diagnosis (talk to lab about sample type, collection, storage, and transport; consider in-country capacity and overseas reference laboratory options) • It is not necessary to confirm all cases

Outbreak Step	Overview of step
3. Make an outbreak case definition	**A standard outbreak case definition** • Allows everyone to include and exclude cases in a comparable way • Should be easily understood and be able to be used by everyone involved in the investigation • May be based on previous case definitions but should be adapted to suit each specific outbreak **Need to find a balance between being** • Too broad (too **sensitive**) – will "screen in" too many false cases • Too narrow (too **specific**) – will "screen out" real cases • Different to surveillance case definitions **A case definition should include elements of:** • Clinical features e.g., Fever, diarrhoea, +/- laboratory confirmation • Person: e.g., age, gender • Place: e.g., village, island, school • Time period: e.g., onset date **May have different categories based on level of certainty** • Confirmed, probable, possible • May incorporate laboratory results **Case definition may change over time as new information is learned in the outbreak** • Start with a **sensitive** case definition (i.e., broad) and refine to be more **specific** as the outbreak develops • Include more laboratory information • Refine based on time, place or person criteria
4. Find cases and find information	**Important information to collect:** • Identifying information: name, date of birth, address, phone number • Demographic information: age (or calculate from DOB), sex, occupation • Clinical details: date of onset, symptom details, severity of illness, hospitalisation • Risk factor information: exposure to known infectious agents or reservoir (e.g., food, travel, gatherings, vaccinations) Use a standardised questionnaire and/or adapt standard case investigation forms as needed. Record information in a 'linelist' (Step 5). **Actively find additional cases:** • Existing surveillance systems • Ask front line staff what they are seeing • Implement passive or active surveillance at health facilities • Conduct site visits if possible • Talk to cases, ask about illness amongst close contacts • Listen to rumours • Issue a public alert encouraging sick people to identify themselves

Outbreak Step	Overview of step
5. Make a linelist	A **linelist** is a table where each ‘row’ contains information about one case (example below) Commonly created in Microsoft Excel, but can be done on paper A linelist makes it easy to describe and interpret your data (Step 6)

ID	First Name	Last Name	Sex	Age (Years)	Village	Province	Date of Presentation	Date of Symptom Onset	Sign / Symptom
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									

Outbreak Step	Overview of step
6. Describe and interpret data	From your linelist, you can start to describe the outbreak (person, place, time) and generate ideas about what is causing it (why, how). You can describe and interpret the data in several ways: **Time** • Develop an **epidemic curve** (epi-curve) • Use illness onset date (or closest date to illness onset) • Date of testing can be considered if no date of symptom onset available **Place** • Village, workplace, exposure(s) of interest • Map cases **Person** • Describe age and sex of cases (minimum age, maximum age, average age, median age, % male, % female) • Describe clinical features (% vomiting, % fever, % hospitalised, % died, etc) After describing the data, you now need to interpret it and to generate ideas as to what is happening (hypothesis generation). For each of the following questions, it is important to consider whether the hypothesis makes sense/is plausible. • What pathogen is causing the outbreak? • How is it spreading? • Likely source? • Who is most at risk? • Why are some people particularly at risk? • Could the outbreak continue to spread? Identify further actions to better understand disease outbreak • Collect additional information from cases • Collect clinical and/or environmental specimens for laboratory testing • Conduct an analytic study (e.g., case-control or cohort study)

Outbreak Step	Overview of step
7. Implement control measures	Act early to minimise impact; do not wait until the end of the investigation. Control measures are often implemented at the same time as the above steps. Use general measures initially and adapt these according to new information and how the outbreak evolves. Measures may include: • Control source: chlorination of water, withdrawal of food product, isolate cases etc • Limit transmission: hygiene measures, mosquito nets, etc • Reduce susceptibility of potential cases (hosts): vaccination, etc
8. Communicate findings	The way that you communicate findings will depend on your target audience. **Sit-reps (situation reports)** • Mainly for senior health staff and government departments • Frequency depends on situation (from none to twice daily) • Length – 1 or 2 pages, maximum A sit-rep will include: • Date / Time • Current outbreak description: Descriptive epi/laboratory results • Response activities underway / planned • Resources needs **Outbreak report** • Mainly for internal health department records • Completed at the conclusion of the outbreak An outbreak report will include: • All items in a sit-rep • Response team and assigned roles • Descriptive/analytic analysis of the outbreak • All response activities undertaken • An epidemic curve as appropriate • Public health recommendations • Conclusions and lessons learned **Timely and targeted public risk communication can:** • Help slow, stop, or prevent outbreaks • Help people to make informed decisions about how to protect themselves • Build public trust in health authorities • Help overcome fear and anxiety • Reduce the economic, social, and political impact of an outbreak **Public risk communication should always** • Be honest and factually correct • Be easily understood • Acknowledge uncertainty and concerns • Avoid excessive reassurance • If in doubt, prepare people for the worst **Common mistakes in public risk communication:** • Waiting until you have all the answers • Withholding bad news • Not telling people what to expect and how to act • Not listening to the public and to rumours

Common Source Outbreak

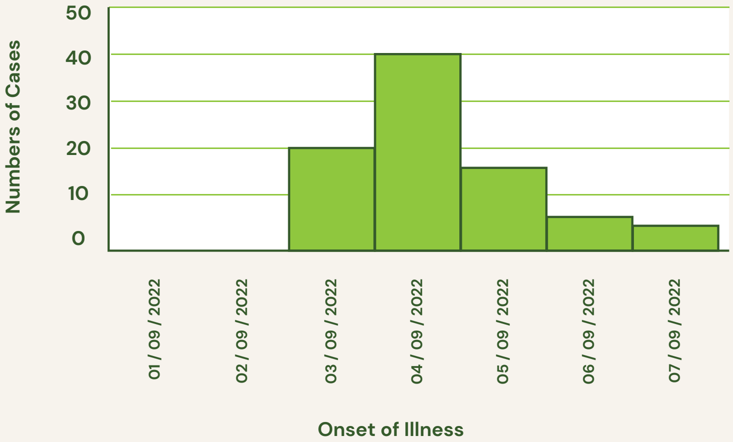
A common-source outbreak is one in which a group of persons are all exposed to an infectious agent or a toxin from the same source. There are three types of common source outbreaks:

Point source outbreak

A type of common source outbreak where the group is exposed over a relatively brief period of time, so that all cases occur within one incubation period.

> **Example:** contaminated food source at a dinner party with 100 guests and 65 guests become cases.

An epicurve in a point source outbreak will have a sharp increase followed by a slower decrease over time. The majority of cases occur within one incubation period.

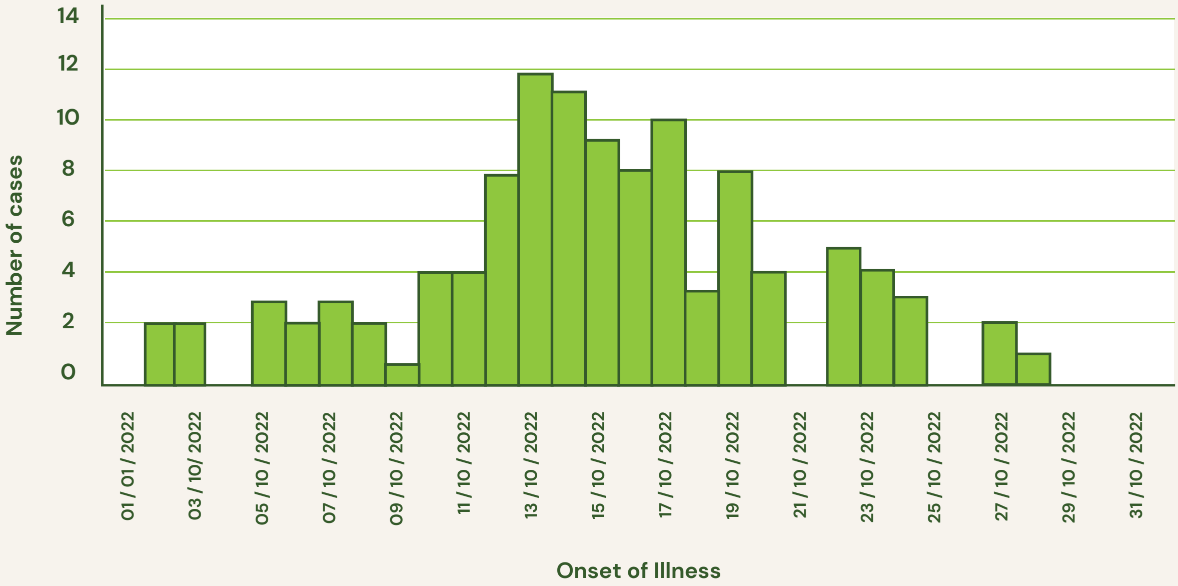


Continuous common source outbreak

A type of common source outbreak where the group is exposed to the source over time, with cases occurring longer than the span of a single incubation period.

> **Example:** contaminated water source to a village where access is not restricted, and people continue to drink the contaminated water.

An epicurve in a continuous common source outbreak will have multiple peaks and declines, with peaks getting larger over time before starting to decrease. There are cases as long as people remain at risk of the common source/the common source remains accessible. Cases are recorded over more than one incubation period.

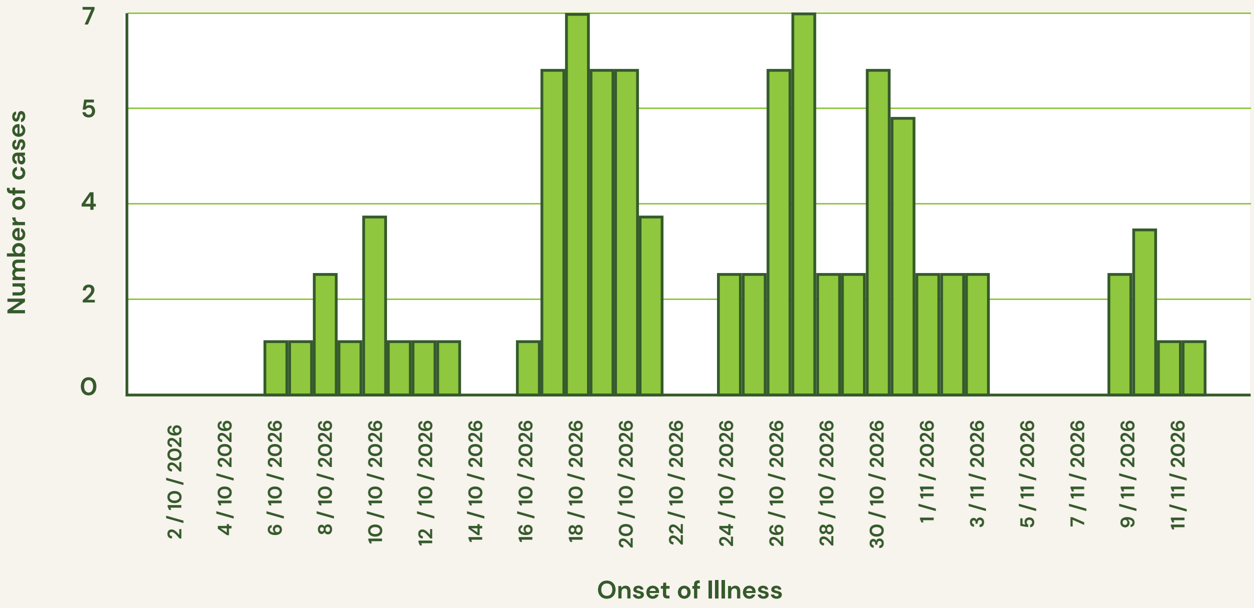


Intermittent common source outbreak

A type of common source outbreak where people are exposed intermittently over time; sometimes they are exposed, and sometimes not.

> **Example:** If a restaurant has five food handlers and one is positive for Hepatitis A, we would expect the number of Hepatitis A cases to reflect when they were exposed to the infected food handler, and not to the other food handlers.

An epicurve in an intermittent common source outbreak will show rapid increases and decreases, reflecting the presence of the source. The length of time between peaks is not related to incubation period, but is instead related to presence of source.

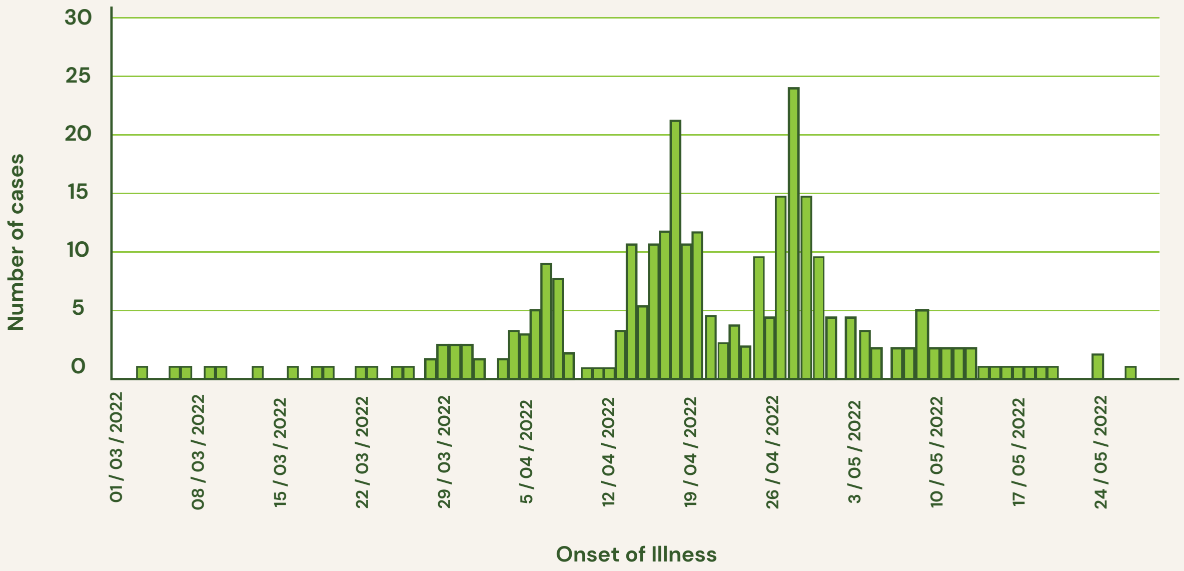


Propagated outbreak

An outbreak that does not have a common source, but instead spreads from person to person.

> **Example:** Measles, where each infected person can infect up to 18 other non-immune people, who can then each infect up to another 18 other non-immune people.

An epicurve in a propagated outbreak usually has a number of irregular peaks reflecting the generations of infection, with peaks separated by approximately one incubation period.



Cluster

A group of cases with a similar disease or set of symptoms over a specified time and in a defined area that are suspected to be greater than the expected number of cases. A cluster is often used to describe a situation prior to a formal outbreak being declared. The difference between a cluster and outbreak is often not clear and they are often used interchangeably amongst the public and media.

> **Example:** Syndromic surveillance may identify a cluster of influenza-like-illness (ILI) cases in a particular village

Contact

Someone who has been exposed to a source of an infection or pathogen.

> **Example:** A person who has spent time with a confirmed COVID-19 case is known as a contact. It means that they were exposed to a confirmed case while the case was infectious. They are not currently a case but have the potential to become a case.

Linelist

A linelist is a way of organising data into a table where each 'row' contains information about one case.
The variables (e.g., Age, sex, clinical characteristics) about each case are in the columns. It is important there is only one row per person and one column per variable (i.e., do not combine multiple symptoms in one column as this cannot be easily analysed). Linelists are commonly created in Excel, but can also be done on paper. A linelist makes it easy to describe and interpret your data.

An example of a linelist is below.

ID	First Name	Last Name	Sex	Age (Years)	Village	Province	Date of illness onset	Nausea	Vomiting	Laboratory confirmation
1	Name	Name	M	36	Village X	Province A	4/12/2022	Yes	Yes	Yes
2	Name	Name	M	42	Village X	Province A	4/12/2022	Yes	Yes	Yes
3	Name	Name	M	63	Village Y	Province A	2/12/2022	Yes	Yes	Yes
4	Name	Name	F	38	Village Y	Province A	1/12/2022	No	Yes	Yes
5	Name	Name	F	25	Village X	Province A	5/12/2022	Yes	Yes	No
6	Name	Name	M	34	Village Y	Province A	7/12/2022	Yes	No	Yes
7	Name	Name	F	43	Village Y	Province A	2/12/2022	No	Yes	Yes
8	Name	Name	M	52	Village X	Province A	3/12/2022	Yes	Yes	Yes
9	Name	Name	F	51	Village X	Province A	5/12/2022	No	No	No
10	Name	Name	M	40	Village Y	Province A	4/12/2022	Yes	Yes	No

Endemic

The constant presence of a disease or infectious agent within a given geographic area or population group.

> **Example:** Malaria is endemic in some Pacific Island nations, in parts of Africa South of the Sahara, parts of Asia, and parts of South and Central America.

Epidemic

The occurrence of more cases of disease than expected in a given area or among a specific group of people over a particular period of time. The increase is often higher than an outbreak or over a longer time period.

> **Example:** Whilst dengue fever can be endemic to a country, there may be a dengue fever epidemic when cases are increased over a long period of time in a particular area.

Pandemic

An epidemic occurring worldwide, or over a very wide area, crossing international boundaries (multiple continents) and usually affecting a large number of people.

> **Example:** COVID-19 pandemic as the virus is affecting large numbers of people on multiple continents.

Sensitivity and specificity ★

Sensitivity and specificity are measures of how well a test identifies a person as either having or not having a disease (or health condition under investigation).

Sensitivity

Sensitivity is the probability a test will return a positive result when the person being tested **has** the condition (true positive). High sensitivity means there are fewer false negatives and therefore less cases are missed. This is important for serious diseases that may require an immediate public health response or for rare diseases.

Specificity

Specificity is the probability a test will show a negative result when the person being tested **does not have** the condition (true negative). High specificity means there are fewer false positives. This is important for diseases that could cause emotional, physical, or social (such as stigma) harm if falsely identified as positive.

It is usually a balance between the two; having higher sensitivity will reduce specificity and the other way around. Each health condition will have different priorities for sensitivity or specificity.

Sensitive or specific case definitions

The principles of sensitivity and specificity also apply to case definitions.

A sensitive case definition aims to identify as many true cases as possible. A sensitive case definition will be broad to identify every case. While it will help detect many cases, it may count people as cases when they don't have the disease.

> **Example:**
>
> Three or more loose stools in a 24-hour period
> AND
> Resident or visitor of Province B
> AND
> Symptom onset on or after January 15, 2022.

A specific case definition aims to correctly exclude noncases. A specific case definition will be narrow to be certain that someone counted as a case really has the disease. A specific case definition may miss counting some people who have the disease.

> **Example:**
>
> Laboratory confirmed case of *E.coli*
> AND
> Resident or visitor of Province B
> AND
> Symptom onset on or after January 15, 2022.

In the early stages of an outbreak, it is important to identify as many cases as possible and a sensitive case definition may be useful. As the outbreak continues and more information is available, case definitions can become narrower (more specific). For example, once a pathogen is identified, the case definition could include laboratory confirmation or more specific clinical symptoms.

Public Health Data

Quantitative data

Data that can be counted. Numbers and classifications that describe an ‘objective reality’. Usually measured through counting or measuring things (e.g., the number of cases of a particular disease). Used to explore cause and effect relationships (the “what”).

There are different types of quantitative data:

CATEGORICAL DATA

Nominal data

Variables with no order or ranking sequence, e.g., Province, Tribe

Ordinal data

Variables with an ordered series, e.g., educational level, age group

Binary data

Variables with only two options, e.g., True/False, Pass/Fail, Yes/No

NUMERICAL DATA

Discrete data

Can only be certain values (e.g., whole numbers) such as counts (e.g. number of students, number of live births)

Continuous data

Can be any value from negative values to zero to infinity, often measured data (e.g., weight, height, length, time)

Qualitative data

Data that are stories and pictures. Stories or words that describe experiences or feelings. Usually collected through interviews, group discussions, observation, and documents.

Used to explore values, attitudes, opinions, feelings, and behaviour (the “why”).

Demographic data

Demographic data are the ‘person’ characteristics of descriptive epidemiology, such as age, sex, occupation, ethnicity.

Variable

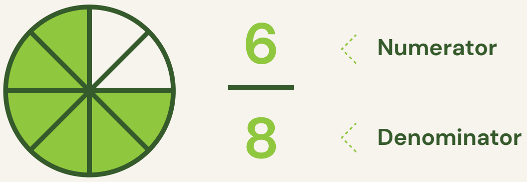
Any characteristic or attribute that can be measured, such as age, sex, village, laboratory diagnosis, blood sugar level.

Numerator

The upper portion of a fraction (number above the line). In epidemiology, the numerator is often the number of people with the disease who meet the case definition.

Denominator

The lower portion of a fraction (number below the line) used to calculate a rate or ratio. In epidemiology, the denominator is often the population at risk.



Example: If eight people sat at a table and ate the food – this is the number of people (population or denominator) at risk from getting sick. If six people got sick (numerator) then 6/8 people became unwell, which we can also express as 3/4.

Median

A measure of central tendency which divides a set of data into two equal parts, identifying the 'middle value' in a set of observations.

1, 3, 5, 7, 12, 18, 25

The Middle Number = Median

Example 1:
What is the median of these numbers?

8, 5, 14

Put the numbers in order:
5, 8, 14

The middle number is 8, **so the median is 8.**

Example 2: What is the median of these numbers?

23, 12, 4, 5, 45, 22, 2, 4, 7, 7, 15, 26

Put the numbers in order:
2, 4, 4, 5, 7, 7, 12, 15, 22, 23, 26, 45

There is an even number of values (n=12) so there is no middle number. In this case, our median is between the two middle numbers (7 and 12) and we need to find the average of these numbers.

Add the two middle numbers and divide by 2 to find the median: **7 + 12 = 19 ÷ 2 = 9.5**

The median value of these numbers is **9.5**

Mean

A measure of central location (tendency) commonly called the average. It is calculated by adding together all the individual values in a set of observations and dividing by the number of observations.

Example 1: What is the Mean of these numbers?

8, 5, 14

Add the numbers: **8 + 5 + 14 = 27**

Divide the total by the number of observations (there are 3 observations in this example): **27 ÷ 3= 9**

The mean is 9

Example 2: What is the Mean of these numbers?

23, 12, 4, 5, 45, 22, 2, 4, 7, 7, 15, 26

Adding all of the above numbers: **172**

Divide the total by the number of observations (there are 12 observations in this example):
172 ÷ 12= 14.3

The mean is 14.3

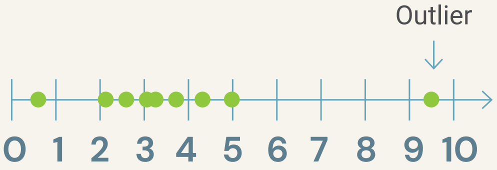
The mean value can be significantly affected by values that are much smaller or larger than the rest of the data (outliers) or if the data are unevenly distributed over the range of the data (skewed). The mean value may be misleading in these circumstances. The median value is much less affected by outliers and skewed data and is usually the better option in these circumstances.

Outlier

Outliers are data points that lie far outside the majority of the values.

Outliers can change the **mean** a lot, so if a dataset has outliers it is better to use the **median** as the measure of central tendency. It is important to check the quality of your data, including of outliers.

Example: in a measles outbreak, there are 22 cases that range from 2 to 14 years of age and there is one case that is 108 years old on the linelist. Here, the outlier would be the person who is 108 years old. This could be a data entry error of the age as we do not expect to find many 108-year-old people in the community, and we also expect older people to be immune to measles. We need to check our data against what we know about the disease and population.



Example:

What is the mean and median of these numbers?

3, 78, 4, 5, 7

The mean = **3 + 78 + 4 + 5 + 7 = 97 ÷ 5 = 19.4**

The median is the middle value, after putting the numbers in order:

3, 4, 5, 7, 78

The median value of these numbers is **5.**

The big difference between the mean and median means there is an outlier in the dataset. In this example, the outlier is **78**.

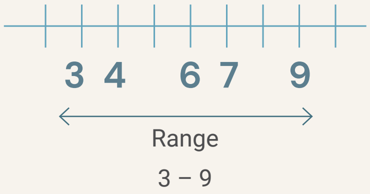
You would need to check that 78 is correct. If it is a valid response, you need to report your data using the median and range.

Mode

The most common value in a set of observations (the number that appears the most times).

Range

In statistics, the difference between the largest and smallest values in a distribution. In common use, the span of values from smallest to largest.



Example: What is the range of these numbers?

3, 1, 7, 5, 9, 4

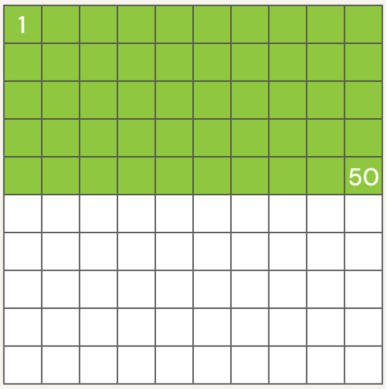
Put the numbers in order: **1, 3, 4, 5, 7, 9**

The range shows the minimum number and the maximum number.

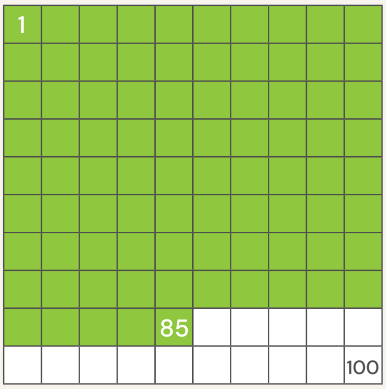
The range = **1 – 9**

Proportion

A comparison of a part to the whole. The numerator (the part) is included in the denominator (the whole). It can be expressed as a fraction (1/5, where numerator = 1, and denominator = 5), a decimal (0.2), or as a percentage (20%).



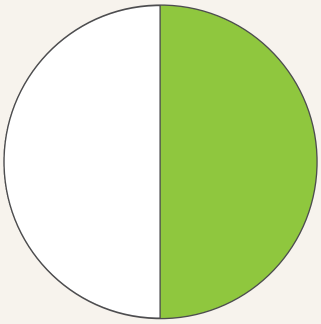
50% means 50 per 100 (50% of this box is green)



85% means 85 per 100 (85% of this box is green)

A percent can be expressed as a decimal or a fraction

A half can be written



As a percentage	50%
As a decimal	0.5
As a fraction	½

In the fraction example above, the numerator = 1 and the denominator = 2.

Calculating Risk or Measures of Association

Attack rate

A measure of the frequency of **new** cases of disease during a specified time, often during an outbreak. It is calculated as the number of cases in the outbreak (numerator) divided by the number of people in the population at risk during the outbreak (denominator). The result is expressed as a percentage (%).

$$\frac{\text{NUMBER OF CASES IN THE OUTBREAK}}{\text{NUMBER OF PEOPLE IN THE POPULATION AT RISK DURING THE OUTBREAK}}$$

Example: During a large measles outbreak in a village of 300 people (denominator) there were 60 people (numerator) with suspected measles.

Attack rate =

$$\frac{\text{NUMBER OF NEW CASES (60)}}{\text{POPULATION (300)}}$$

= 0.2

To make a percentage, multiply by 100 = 20%

The attack rate of measles in this outbreak is 20%.

Incidence proportion

A measure of the frequency of **new** cases of a particular disease in a population over a period of time. It is calculated as the number of new cases of a disease in a specified time (numerator) divided by the number of people in the population at risk at that time (denominator). The population at risk does not include existing cases of the disease.

It is usually expressed in terms of a common denominator (e.g., per 100,000 population) to allow the comparison of different populations.

$$\frac{\text{NUMBER OF NEW CASES OCCURRING DURING A TIME PERIOD}}{\text{POPULATION AT RISK}}$$

Example: There are 5 people newly diagnosed with TB during 2020 in Province X that has a population of 10,000 people.

Incidence rate =

$$\frac{\text{NUMBER OF NEW CASES (5)}}{\text{POPULATION (10,000)}}$$

= 0.0005

To express per 100,000 population, multiply 0.0005 by 100,000

= 50 newly diagnosed TB cases per 100,000

Prevalence

The **proportion** of a population who have a specific disease or condition in a given time period regardless of when they first developed the condition. Prevalence is a particularly useful measure for chronic communicable diseases, or diseases of longer duration (such as non-communicable diseases-NCDs) where we want to measure the proportion of a population who has a condition at any one time.

While we can also measure the **incidence** of NCDs, if we only calculated incidence of diabetes (or other chronic disease) in a village, for example, we would only count the newly-diagnosed cases in the village during that time period, so we would miss the people who were diagnosed many years ago and continue to live with diabetes. The prevalence shows us the total number of people living with diabetes in that time period.

It is calculated as the number of existing cases of a disease in a population on a particular date (numerator), divided by the number of people in the population on that date (denominator).Prevalence may be reported as a percentage (5%, or 5 people out of 100), or as the number of cases per 100,000 population.

$$\frac{\text{NUMBER OF EXISTING CASES OF A DISEASE AT ONE TIME POINT}}{\text{NUMBER OF PEOPLE IN THE SAMPLE POPULATION AT THE SAME TIMEPOINT}}$$

> **Example:** At the end of 2020, it was found that there were 40 people living with TB in Province X (population 10,000 people).

$$\frac{\text{NUMBER OF EXISTING CASES (40)}}{\text{POPULATION (10,000)}}$$

This shows there are 0.004 cases of TB per person which is confusing when communicating this.

To communicate this as the number of cases per 100,000 people we:

multiply 0.004 by 100,000

= 400 TB cases per 100,000 population

This can also be expressed as a percentage. To express this as a percentage of the population living with TB, we take the original calculation and multiple by 100.

$$\frac{\text{NUMBER OF EXISTING CASES (40)}}{\text{POPULATION (10,000)}}$$

Then to make this a percentage, we multiply 0.004 by 100 = 0.4%

Expressing a prevalence per 100,000 population does not mean there are really 100,000 people. It helps us to meaningfully compare rates across population groups as though the populations were the same.

> **Example:** There are 24 cases of TB in Province A, which has a population of 25,000 people. There are 30 cases of TB in Province B which has a population of 80,000 people. We would like to compare what the burden of TB is like in each of these two places. To do this, we consider the number of cases in each place as though they both have a population of 100,000 people.

Province A:

$$\frac{\text{NUMBER OF EXISTING CASES (24)}}{\text{POPULATION (25,000)}}$$

= 0.00096

To express per 100,000 population, multiply 0.00096 by 100,000

= 96 TB cases per 100,000

Province B:

$$\frac{\text{NUMBER OF EXISTING CASES (30)}}{\text{POPULATION (80,000)}}$$

= 0.000375

To express per 100,000 population, multiply 0.000375 by 100,000

= 38 TB cases per 100,000

If we simply look at the number of cases for both Province A (n=24) and Province B (n=30), Province B has a larger number of cases. Because these two geographic areas have a different number of people in their populations, however, we cannot truly compare these numbers meaningfully.

When we calculate the prevalence per 100,000 population (i.e., We calculate as though both places have a population of 100,000 people), then we can see that Province A has a much higher prevalence (96 per 100,000) compared with Province B (38 per 100,000).

Two-by-two table (contingency table)

A two-by-two table is a two-variable table in which the data of one variable are cross-tabulated with another variable. It is also referred to as a contingency table. The table is used to investigate the relationship between two variables in a study. Usually, the 'dependent' variable is displayed at the top of the table in columns (also called the disease or outcome) and the independent variables are represented across the table in rows (also called the 'exposure' or 'treatment').

For calculations of measures of association such as odds ratios and risk ratios, the cells are labelled A-D as shown below:

		Disease or Outcome		
		With Disease	Without Disease	TOTAL
Exposure or treatment or intervention	Had exposure	A	B	A + B
	Did not have exposure	C	D	C + D
	Total	A + C	B + D	A + B + C + D

A and B (shaded green) includes counts of people who had exposure to the item being studied
C and D (shaded orange) includes counts of people who were not exposed to the item being studied

		Disease or Outcome		
		With Disease	Without Disease	TOTAL
Exposure or treatment or intervention	Had exposure	A	B	Total exposed (A + B)
	Did not have exposure	C	D	Total not exposed (C + D)
	Total	A + C	B + D	A + B + C + D

A and C (shaded green) includes counts of people who have the disease or outcome being studied (cases)
B and D (shaded orange) includes counts of people who do not have the disease or outcome

		Disease or Outcome		
		With Disease	Without Disease	TOTAL
Exposure or treatment or intervention	Had exposure	A	B	Total exposed (A + B)
	Did not have exposure	C	D	Total not exposed (C + D)
	Total	Total with disease (A + C)	Total without disease (B + D)	A + B + C + D

Relative Risk/Risk Ratio

Relative risk or Risk Ratio (RR) compares the **risk** of a health event in one group with the **risk** of the event in another group. Often the groups differ in their exposure to a certain item (e.g., did or did not eat a food item) or they could differ by other factors (e.g., males compared with females).

Relative risk is a more robust measure of association than, for example, odds ratio but can only be used in a study in which groups are selected based on a certain exposure and the total number exposed are known (e.g., cohort study).

To calculate the relative risk, the attack rate of the disease in both the exposed and unexposed groups needs to be calculated.

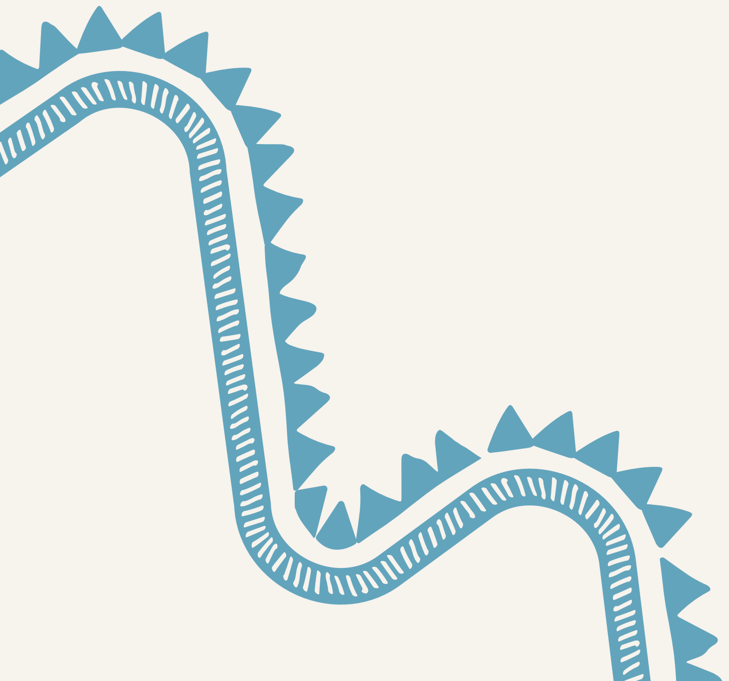
$$\text{Relative risk} = \frac{\text{RISK (ATTACK RATE) OF DISEASE IN THE EXPOSED}}{\text{RISK (ATTACK RATE) OF DISEASE IN THE UNEXPOSED}}$$

Using a 2x2 table this equates to:

$$\text{ATTACK RATE IN THE EXPOSED} = \frac{A}{A + B}$$

$$\text{ATTACK RATE IN THE UNEXPOSED} = \frac{C}{C + D}$$

$$\text{RELATIVE RISK} = \frac{(A / (A + B))}{(C / (C + D))}$$



Interpreting relative risk:

- If RR = 1 there is no difference in risk between the two groups being studied
- If RR <1 there is a decreased risk for the exposed group (they are less likely to get the disease than the unexposed group), indicating that the exposure potentially offered some protection against the outcome being studied
- If RR >1 there is an increased risk for exposed group (they are more likely to get the disease than the unexposed group)

Example: Field epidemiologists were alerted to a cluster of people with gastroenteritis symptoms who all attended a workshop dinner. The epidemiologists wanted to understand if the sickness was related to chicken consumed at the dinner.

They obtained the full guest list for the dinner, which had 40 people. Because they were able to identify all people who attended the dinner, and therefore able to determine exposure status for everyone, they could calculate a relative risk.

16 of these 40 people developed gastroenteritis over the 24 hours following the workshop dinner. Of the 16 people with gastroenteritis, 14 had eaten chicken for dinner and 2 had not eaten chicken.

The **2x2 table** to measure the association between exposure with chicken and outcome (gastroenteritis) is presented below.

		Gastroenteritis		
Consumed Chicken		With Disease	Without Disease	TOTAL
	Had exposure	14	6	20
	Did not have exposure	2	18	20
	Total	16	24	40

In this 2x2, the interpretation of each cell is as follows:

- A = there were 14 people who ate chicken and also had gastroenteritis
- B = there were 6 people who ate chicken but did not have gastroenteritis
- C = there were 2 people who did not eat chicken but did have gastroenteritis
- D = there were 18 people who did not eat chicken and who also did not have gastroenteritis

To calculate the relative risk use the numbers in the 2x2 in the RR formula:

$$\frac{A/(A+B)}{C/(C+D)} = \frac{14/(14+6)}{2/(2+18)} = \frac{14/20}{2/20} = \frac{0.7}{0.1} = 7$$

The relative risk of gastroenteritis amongst those who ate chicken was 7.0.

We would interpret this to mean that those people who attended the workshop dinner who ate chicken were 7 times more likely to develop gastroenteritis than those who attended the workshop dinner and did not eat chicken.

Odds ratio

An odds ratio (OR) is a measure of association between two variables comparing the ratio of **odds** of the event in one group with the **odds** of the event in the other group. It is used in studies where the groups are selected based on the outcome being studied (e.g., cases of a disease, unsupervised deliveries) not the “exposure” or “independent variable” (e.g., time to the clinic), for example case control studies and cross-sectional studies. Because the true denominator (the total number of people exposed) for the exposure is not known, a direct measure of risk is not possible. So, we instead compare the odds of “exposure” between the groups.

Odds ratio (OR) =

$$\frac{\text{ODDS OF EXPOSURE AMONGST CASES}}{\text{ODDS OF EXPOSURE AMONGST CONTROLS}}$$

Using a 2x2 table, this equates to:

$$\frac{A \times D}{B \times C}$$

Interpreting Odds Ratios

- If OR = 1 there is no difference between cases and controls in the odds of the exposure.
- If OR <1 the odds of exposure is less for those who had the disease compared to those who did not
- If OR >1 the odds of exposure is higher for those who had the disease compared to those who did not
- If the condition we are studying is rare, then an OR calculation is a good estimate of RR calculation.

Example: Field epidemiologists wanted to investigate if there was any association between overseas travel and influenza infection.

They asked 645 people who had had influenza and 256 people who did not have influenza (but had had another respiratory illness) if they had travelled overseas in the week before they fell ill.

Amongst those who had had influenza, 134 had travelled overseas.

A total of 750 in the study did not travel overseas.

The 2x2 to calculate the odds ratio for this study would be completed as follows, where the exposure is overseas travel in the week prior to falling ill, and the outcome/disease is influenza infection.

		Disease or Outcome		
		With Disease	Without Disease	TOTAL
Overseas Travel	Had exposure	134	17	151
	Did not have exposure	511	239	750
	Total	645	256	901

In this 2x2, the interpretation of each cell is as follows:

A = there were 134 people who had influenza infection who also travelled overseas in the week prior to becoming ill

B = there were 17 people who did not have influenza who did travel in the week prior to becoming ill

C = there were 511 people who had influenza and did not travel overseas in the week prior to becoming ill

D = there were 239 people who did not have influenza and did not travel

To calculate the odds ratio in this example, we would then calculate:

$$OR = \frac{A \times D}{B \times C} = \frac{134 \times 239}{17 \times 511} = \frac{32026}{8687} = 3.69$$

Odds ratio = 3.69

The interpretation of this result would be: The odds of having travelled overseas in the week prior to falling ill was 3.69 times higher for those who had influenza compared to those who did not have influenza.

Confidence intervals

It is not feasible to study a whole population, so we usually select a sample of that population to study. We understand that by chance, our sample may not perfectly represent the population.

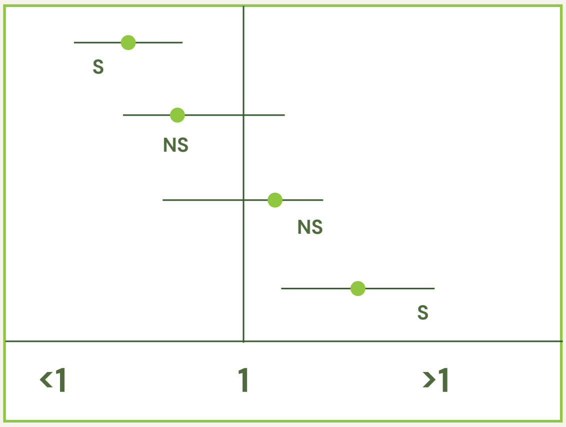
A confidence interval is calculated to understand how well we have estimated the population with our sample group; the interval gives us a range of plausible values within which the true population value is likely to lie.

In field epidemiology, typically we estimate to 95% confidence. The 95% confidence interval gives a range of values either side of our estimate (whether that be an odds ratio, a risk ratio or other measure) within which we can be 95% sure the true population value lies (we are confident that 95 out of 100 times the estimate would fall between the lower and upper intervals).

Interpreting confidence intervals

1. In general, the confidence interval becomes narrower the bigger your sample size and the less variation there is in the population.
2. A narrow confidence interval indicates high levels of certainty about the true value, whilst a wide confidence interval indicates lower levels of certainty.
3. For odds ratios and risk ratios, if the interval crosses 1 then we generally conclude there is no difference between the groups (see diagram).
4. If we are comparing another measure between groups such as the “average distance from the health centre” when the interval crosses zero we generally conclude that there is no difference between the groups.

Example: An early study of the COVID-19 case fatality rate among the 15–44 year age group in March 2020 was estimated to be 0.5% (95% CI: 0.1%–1.3%). The interpretation for this, therefore, is that: for those aged 15 to 44 years, the fatality was 0.5%. It may, however, have been as low as 0.1% or as high as 1.3%.



S = Significant NS = Not =Significant

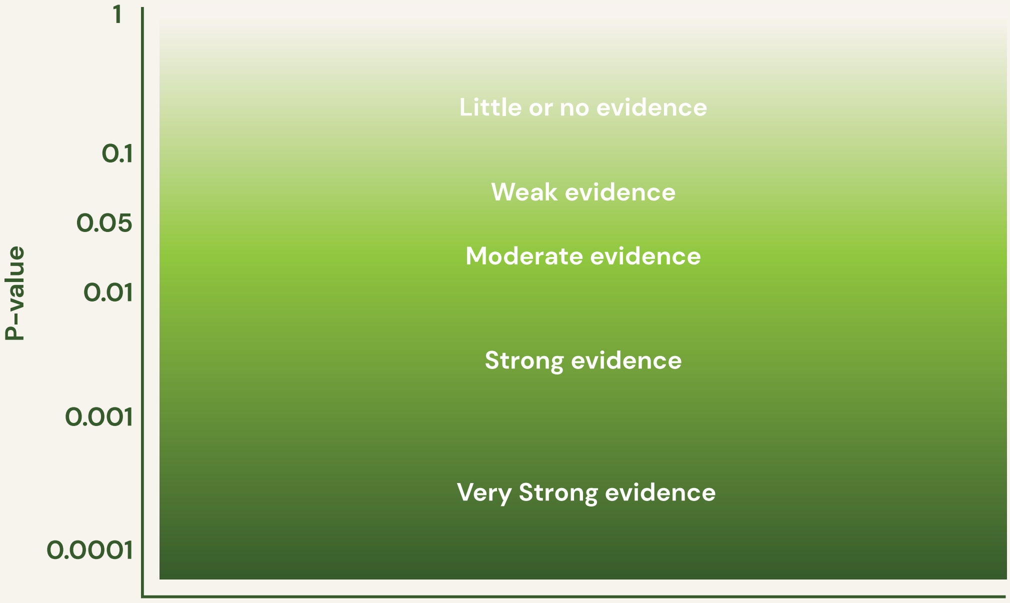
P-values

A p-value is a measure of probability (chance) and ranges between 0 (no chance) and 1 (certainty).

A p-value is the probability that the study estimate would be as large as (or larger than) we observed if the **null hypothesis** was true. The **null hypothesis** is true when:

- There is no actual difference between the two groups being compared *OR*
- There is no association between an exposure and an outcome (i.e., RR or OR = 1), *OR*
- There has been no effect of an intervention on an outcome (in intervention studies)

A p-value shows the strength of evidence against the **null hypothesis** (as illustrated below):



Source: https://theoreticalecology.wordpress.com/2021/12/08/can-p-values-be-interpreted-as-continuous-measures-of-evidence-for-an-effect/

In other words, the p value gives an idea of the strength of evidence of our study estimate whether that's an association between two groups (e.g., exposed and unexposed), an estimate of a population mean, or an estimate effect size of an intervention.

We typically assume that p-values <0.05 are significant, and those >0.05 are not significant, with the lower the p-value indicating the stronger the evidence against the **null hypothesis**. This means if you reproduced the study (with the same conditions) 100 times, and assuming the **null hypothesis** is true, you would see the results only 5 times, or there's only a 5% chance of seeing the results.

Presenting Public Health Data

Tables

Tables are a visual way of presenting and summarising public health data. Tables must include:

- A descriptive title that includes the table number and the 'what', 'where', 'when' details of the data presented
- A clear and concise label for each row and column, including units of measurement (e.g., years)
- Totals for rows and/or columns as appropriate
- All codes, abbreviations, and symbols explained either in the Table or below in a footnote
- Any exclusions from the total dataset explained in a footnote
- Source(s) of information noted

Below is an example of a table with all required components:

Table 1. Reported cases of Malaria by District, Province A, 2020

District	Malaria Cases	
	(n)	(%)
District A	55	16
District B	40	12
District C	50	15
District D	40	12
District E	60	18
District F	50	15
District G	45	13
Total	340	100

Graphs

Graphs are visual ways of presenting data in different formats.

Whilst there are several different kinds of graphs, there are similarities in the way that they are presented. A graph should be able to be interpreted on its own (without needing to read results).

All graphs must include the following elements:

- Clear title that includes the Figure number with 'what', 'where', and 'when' details
- Labels on both both x (horizontal) and y (vertical) axes that includes the unit/s of measurement
- The y axis starts at zero
- A key to explain the different categories
- An explanation of any abbreviations or symbols

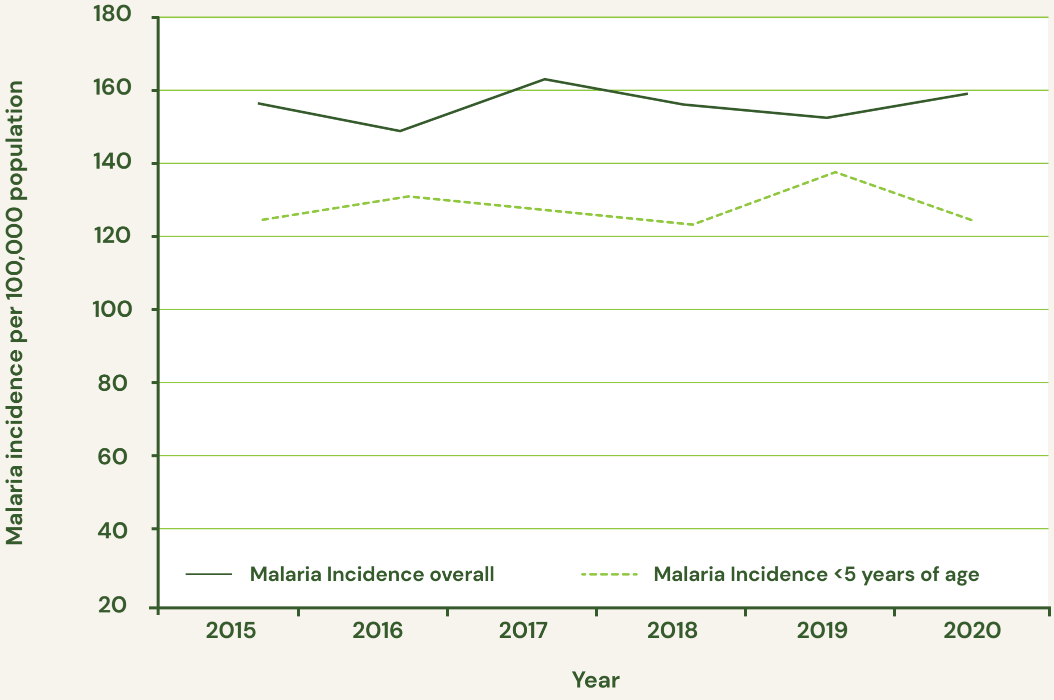
Line graph

A line graph is a visual way to show changes/patterns/trends over time.

Time is represented on the x axis and the number or rate is represented on the y axis. By using multiple lines (with different colours or patterns) we can compare multiple categories on the one line graph.

Below is an example line graph with the key required elements.

Figure 1. Malaria incidence per 100,000 population (overall and under five years of age), Country X, 2015–2020



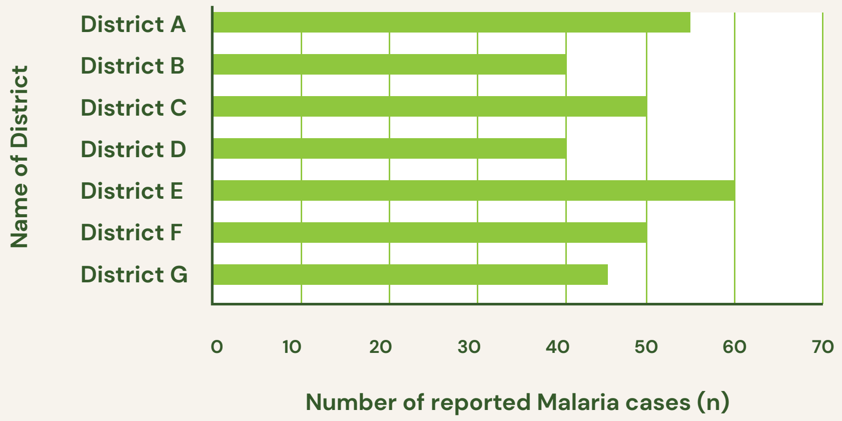
Bar chart

A visual way of displaying categorical data, where each bar represents a different category, such as age group or sex. It helps to visually show a comparison between different categories.

The length of the bar corresponds with the frequency of the category it relates to. Bar charts are presented with the categories on the y axis – that is, the bars will run horizontally. The bars are not touching as they represent separate, distinct categories.

Below is an example bar chart with the key required elements:

Figure 2. Reported cases of Malaria by District, Province A, 2020

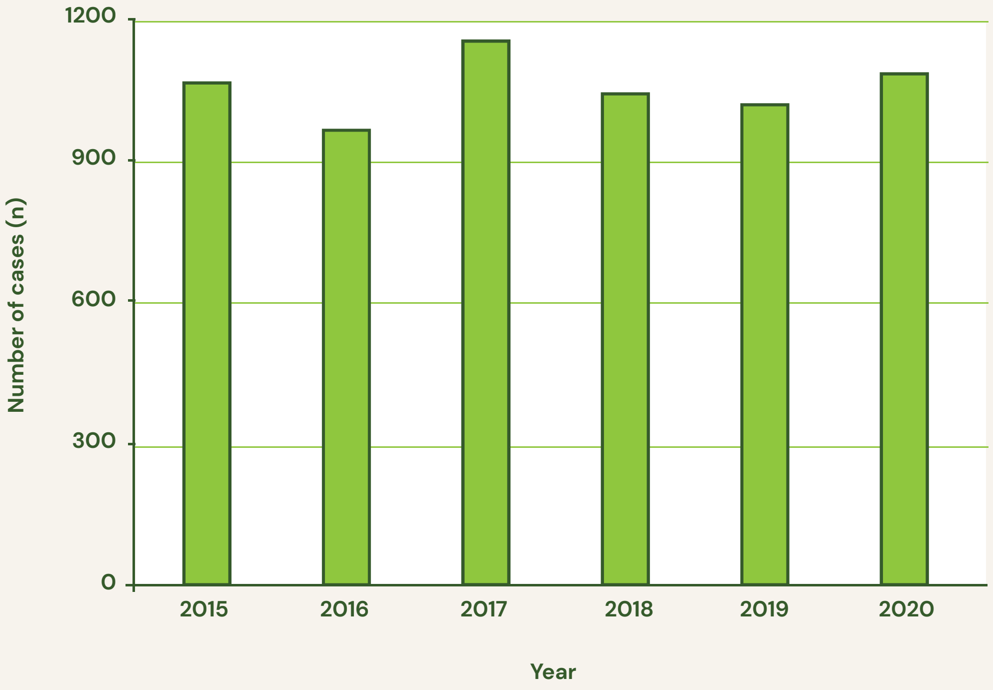


Column chart

A visual display of the size of the different categories of a variable. Each category or value of the variable is represented by a column. The y axis shows the number of cases and the x axis shows the different categories of data that can be presented (bars run vertically). There are gaps between the columns because the y axis represents separate, distinct categories. There are four types of column charts (simple, grouped, stacked, 100%) and the type that you choose to use depends on the data and desired emphasis.

Below is an example column chart with the key required elements.

Figure 3. Malaria cases by year, Country X, 2015–2020

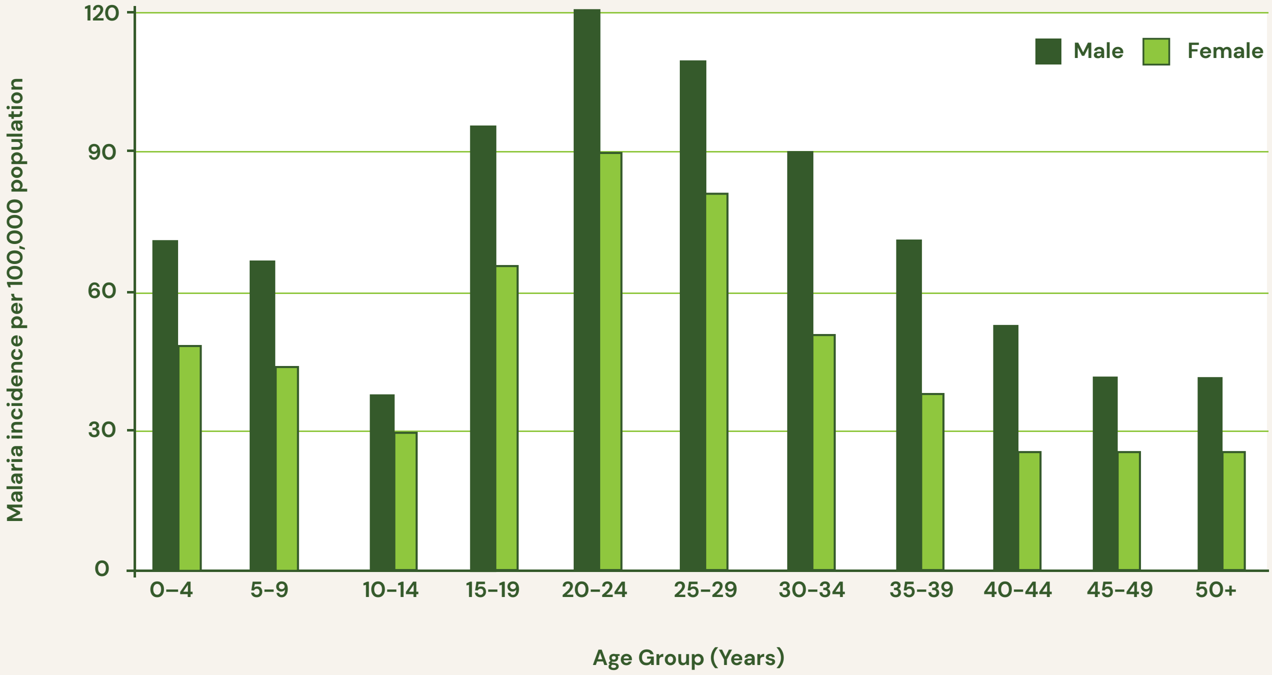


Grouped column chart

A grouped column chart illustrates data from tables with two or three variables. It is useful to display data in a way that enables easy comparison between sub-sets of a group. Within each group, the bars are touching. The bars of separate groups are not joining.

> **Example:** In the below figure, the y axis shows the incidence per 100,000 population and the x axis shows age groups in years. The x axis also shows the breakdown by sex as seen by the different coloured bars and the colour legend. By presenting these data in this grouped column chart, we can compare the incidence by both age and sex at the same time.

Figure 4. Malaria incidence by age group and sex, Country X, 2020

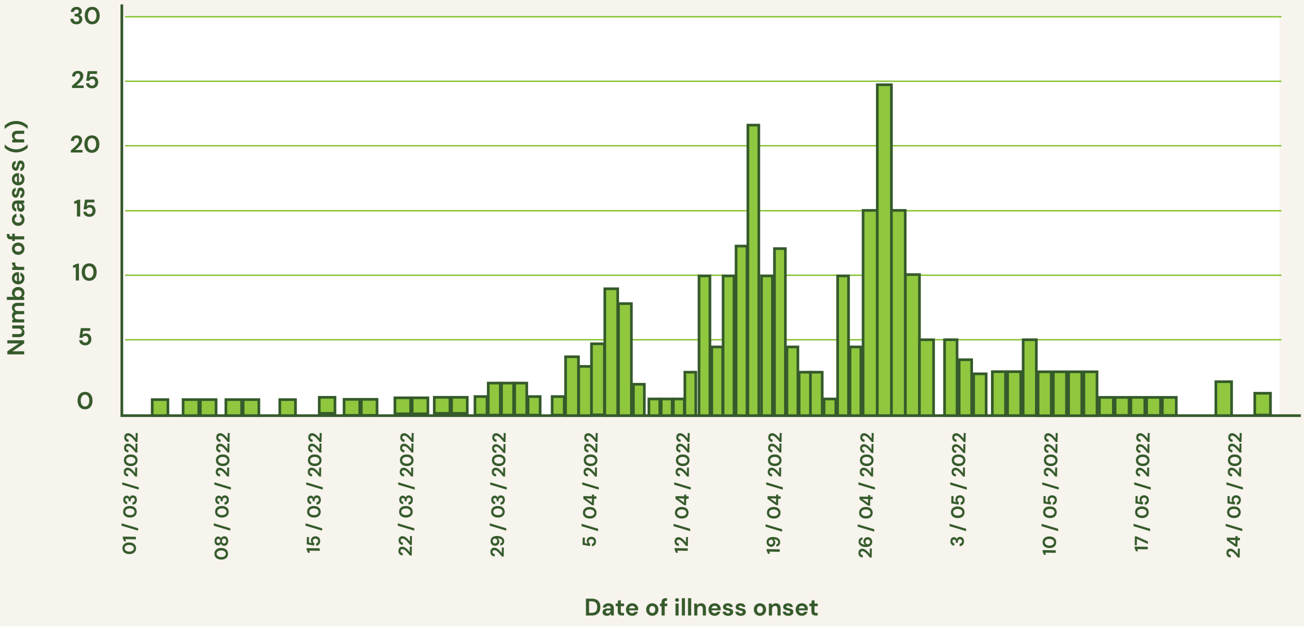


Epidemic curve (epicurve)

A histogram that shows the course of a disease outbreak or epidemic by plotting the number of cases on the y (vertical) axis by date/time of onset on the x (horizontal) axis. Each case (sick person) can only be plotted once on the epi curve; this is the day of onset of symptoms. If there are asymptomatic cases (which means they have no date of symptom onset), the case is entered on the date of testing/presentation.

Epicurves should have whole numbers on the y axis (not decimals); e.g., In increments of 1, 10, 100, etc (depending on how many cases); not 0.5, 1.5 etc. The columns are always next to each other with no space in between. Below is an example of an epicurve with the key required elements.

Figure 5. Measles cases by date of onset, Province A, 2022



Operational Research

Descriptive epidemiology

Descriptive epidemiology is describing the what, when, where, and who of health determinants or conditions in a population. What (clinical features); who (person); where (place); when (time). Descriptive epidemiology can be used for surveillance data, field projects and in outbreak investigations. With descriptive epidemiology we can start to form hypotheses about a possible source of an illness during an outbreak investigation.

Analytic epidemiology

The form of epidemiology that looks to identify and measure associations, test hypotheses, and identify causes. Uses a comparison group (or groups) to confirm or reject an association between risk factors and illness between the groups. Case-control studies and cohort studies are used as analytical studies to test hypotheses about a possible source of illness during an outbreak investigation.

Association

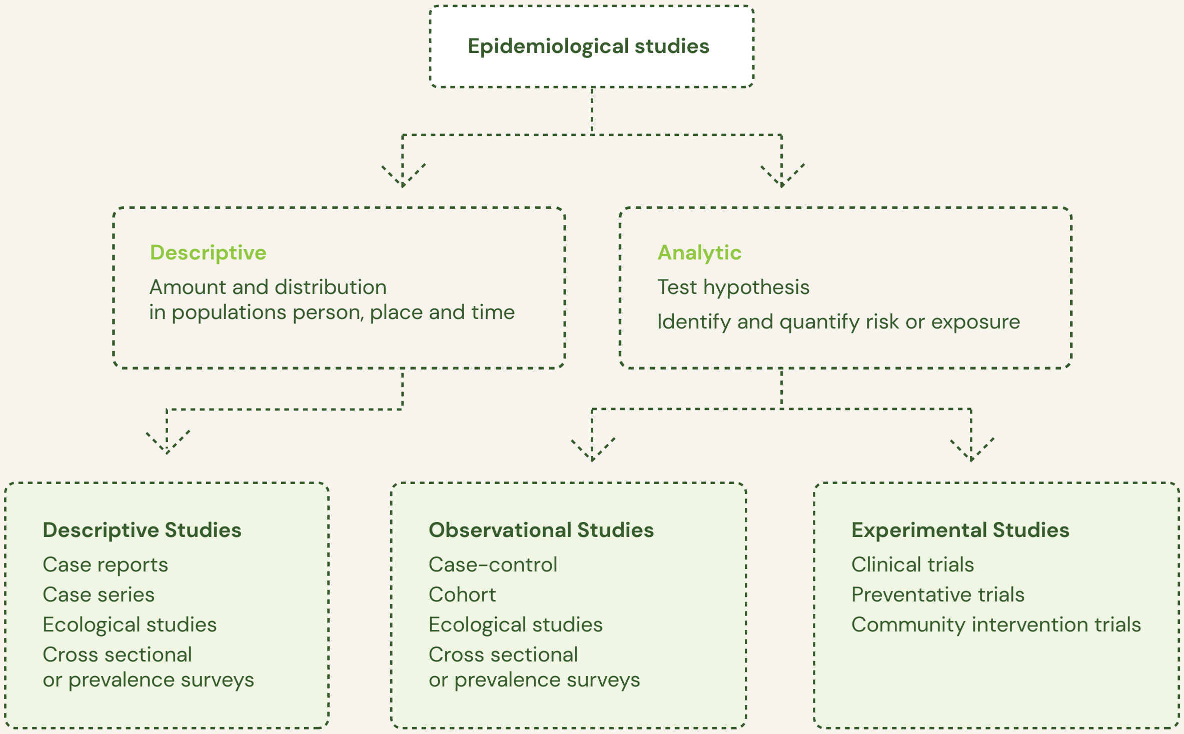
A statistical measure of the relationship between two or more events. In epidemiology, the events we are typically looking at are exposure and outcome (disease).

Experimental study

A study in which the investigator specifies the exposure category for each individual (clinical trial) or community (community trial), then follows the individuals or community to detect the effects of the exposure.

Observational study

Epidemiological study in situations where nature is allowed to take its course. Changes or differences in one characteristic are studied in relation to changes or differences in others, without the intervention of the investigator.



Epidemiological studies

Descriptive
Amount and distribution
in populations person, place and time

Analytic
Test hypothesis
Identify and quantify risk or exposure

Descriptive Studies
Case reports
Case series
Ecological studies
Cross sectional
or prevalence surveys

Observational Studies
Case-control
Cohort
Ecological studies
Cross sectional
or prevalence surveys

Experimental Studies
Clinical trials
Preventative trials
Community intervention trials

Hypothesis

A hypothesis is a proposed explanation or prediction for something. It is based on known information but has not yet been proven to be true. A hypothesis is tested through further study (analysis, experiments, observation) to see whether it is a valid explanation.

Null hypothesis

The null hypothesis assumes that there is no difference or association between the items under study (could be groups, populations, variables etc). The null hypothesis assumes any observed difference is due to chance and not a true association.

Alternative hypothesis

The alternative hypothesis goes against the null hypothesis and predicts that there is a significant difference or association between the items under study.

Example:

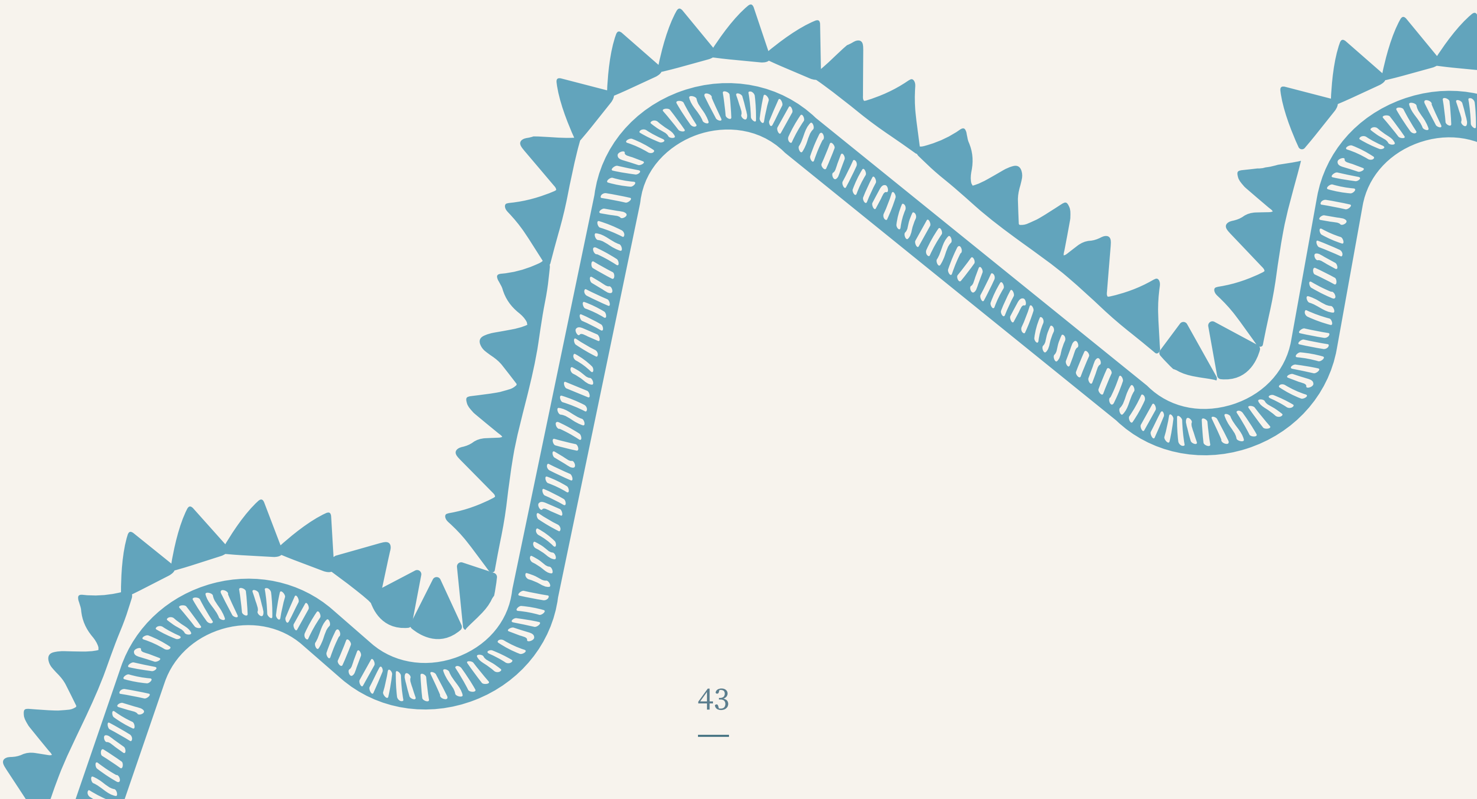
An epidemiologist is studying the relationship between malaria incidence and annual average rainfall. Their hypothesis is that malaria incidence is higher in locations with high rainfall compared with locations with low rainfall.

The null hypothesis would be that there is no significant difference in the incidence of malaria between areas with high rainfall and areas with low rainfall.

The alternative hypothesis would be that the incidence of malaria is significantly higher in areas with high rainfall compared with areas with low rainfall.

The epidemiologist would then design an appropriate study to test these hypotheses.

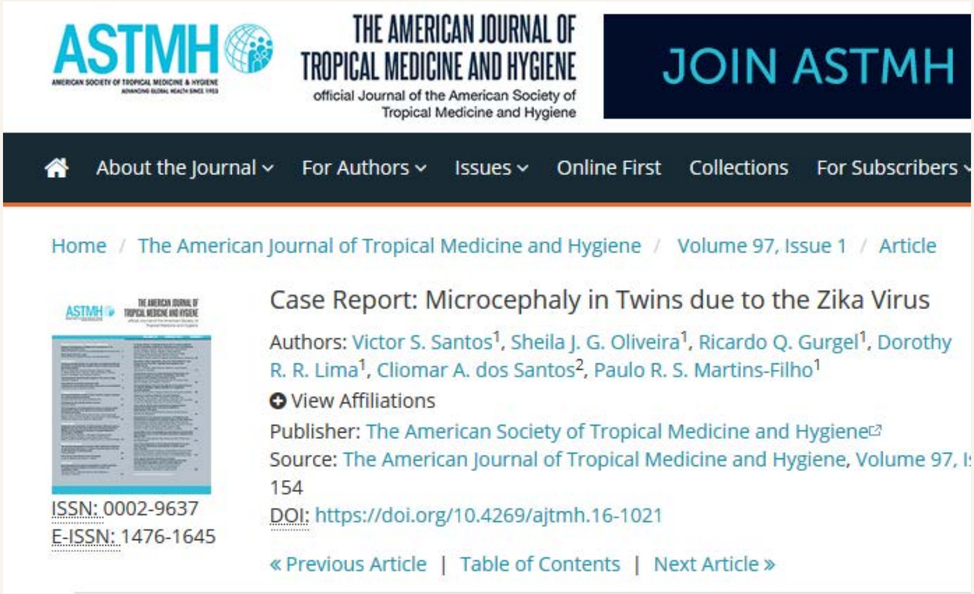
A descriptive study can help generate hypotheses, whereas an analytic study can test them.



Case reports

A type of descriptive epidemiological study that reports individual cases of unusual diseases or outcomes.
Often clinical reports are of rare diseases or rare outcomes. Often used to describe first case of an emerging disease.

> **Example:** When Zika virus began spreading rapidly across Brazil in 2015, case reports were written about individual cases or, in this case, twins who had microcephaly and confirmed Zika virus. In this article, the authors describe in detail the clinical characteristics of these two babies.


ASTMH
AMERICAN SOCIETY OF TROPICAL MEDICINE & HYGIENE
THE AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE
official Journal of the American Society of Tropical Medicine and Hygiene
JOIN ASTMH
About the Journal
For Authors
Issues
Online First
Collections
For Subscribers

Home / The American Journal of Tropical Medicine and Hygiene / Volume 97, Issue 1 / Article

Case Report: Microcephaly in Twins due to the Zika Virus

Authors: Victor S. Santos[1], Sheila J. G. Oliveira[1], Ricardo Q. Gurgel[1], Dorothy R. R. Lima[1], Cliomar A. dos Santos[2], Paulo R. S. Martins-Filho[1]

View Affiliations

Publisher: The American Society of Tropical Medicine and Hygiene

Source: The American Journal of Tropical Medicine and Hygiene, Volume 97, I
154

DOI: https://doi.org/10.4269/ajtmh.16-1021

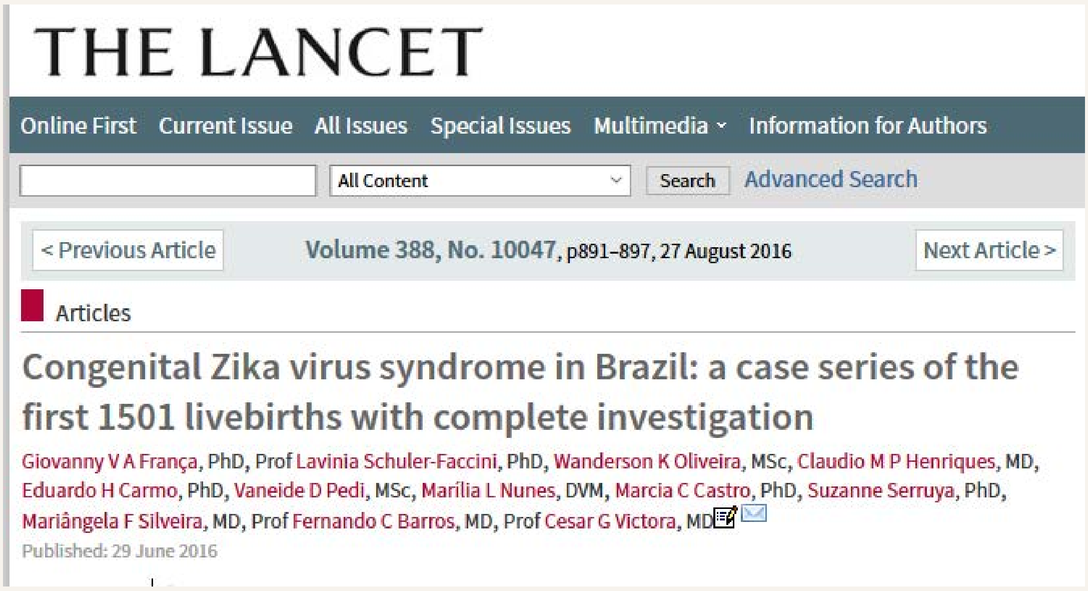
ISSN: 0002-9637
E-ISSN: 1476-1645

« Previous Article | Table of Contents | Next Article »

Case series

A descriptive epidemiological study that is similar to case reports but describes multiple cases.

> **Example:** In the Zika virus outbreak in 2015, there were then enough cases detected to produce case series – in this example below, the first 1501 live births with complete investigation in Brazil. This article describes the clinical characteristics of the first 1501 live births.



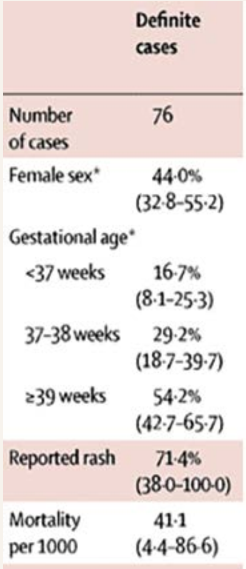
< Previous Article | Volume 388, No. 10047, p891–897, 27 August 2016 | Next Article >

Articles

Congenital Zika virus syndrome in Brazil: a case series of the first 1501 livebirths with complete investigation

Giovanny V A França, PhD, Prof Lavinia Schuler-Faccini, PhD, Wanderson K Oliveira, MSc, Claudio M P Henriques, MD, Eduardo H Carmo, PhD, Vaneide D Pedi, MSc, Marília L Nunes, DVM, Marcia C Castro, PhD, Suzanne Serruya, PhD, Mariângela F Silveira, MD, Prof Fernando C Barros, MD, Prof Cesar G Victora, MD

Published: 29 June 2016



	Definite cases
Number of cases	76
Female sex*	44·0% (32·8–55·2)
Gestational age*	
<37 weeks	16·7% (8·1–25·3)
37–38 weeks	29·2% (18·7–39·7)
≥39 weeks	54·2% (42·7–65·7)
Reported rash	71·4% (38·0–100·0)
Mortality per 1000	41·1 (4·4–86·6)

Cross-sectional study

A cross-sectional study looks at data from a population at one point in time; it is a snapshot of what is happening in this population at this point in time. A cross-sectional study is a type of observational study. It can also be a descriptive study.

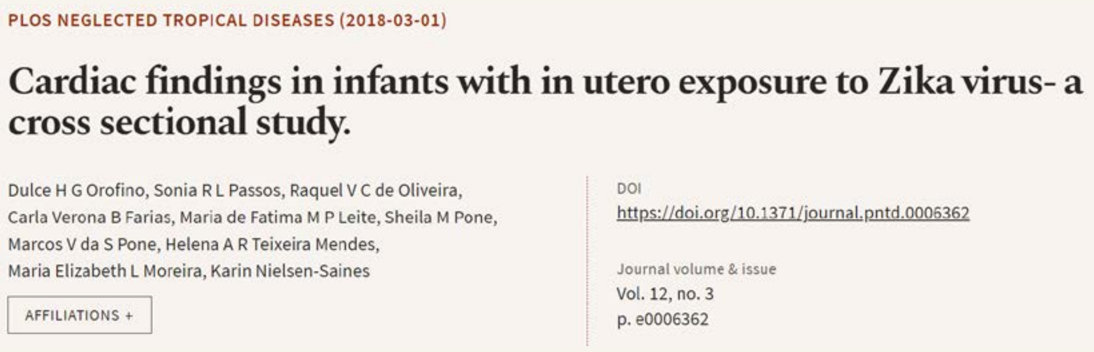
The exposure and outcome are measured at the same time, so it does not examine cause and effect; you cannot use a cross-sectional study to see what causes a disease. Participants are selected based on the variables you want to study. The researcher records all of the data variables present in the population; they do not do anything to try to change these variables. Often called a ‘prevalence study’.

In designing these studies, we identify a research question, specify a population, sample that population, and measure variables of interest. By doing this, we are able to easily estimate the prevalence of diseases within the population.

Advantages and disadvantages of cross-sectional studies include:

Advantages	Disadvantages
• Can study entire populations or a representative sample • Provide estimates of prevalence of all factors measured • Standardized data collection tool • May be quick and inexpensive • Valuable in assessing health status and health care needs of a population • Can be repeated to get trend data	• Difficult to determine which came first, the exposure or the outcome • Not suitable for studying rare or short-lived diseases • Data are affected by survival • Findings may be hard to trust – low response rates or poor recall of subjects

> **Example:** The below cross-sectional study looked at 120 infants who had been exposed to Zika virus in utero to see the proportion of these infants who also had heart issues.



PLOS NEGLECTED TROPICAL DISEASES (2018-03-01)

Cardiac findings in infants with in utero exposure to Zika virus- a cross sectional study.

Dulce H G Orofino, Sonia R L Passos, Raquel V C de Oliveira, Carla Verona B Farias, Maria de Fatima M P Leite, Sheila M Pone, Marcos V da S Pone, Helena A R Teixeira Mendes, Maria Elizabeth L Moreira, Karin Nielsen-Saines

AFFILIATIONS +

DOI
https://doi.org/10.1371/journal.pntd.0006362

Journal volume & issue
Vol. 12, no. 3
p. e0006362

By looking at the variables – cardiac findings and in utero exposure to Zika virus – the study team were able to measure the frequency of cardiac findings amongst those children with in-utero exposure to the virus. They determined that amongst those infants with in-utero exposure, there was higher prevalence of major cardiac defects.

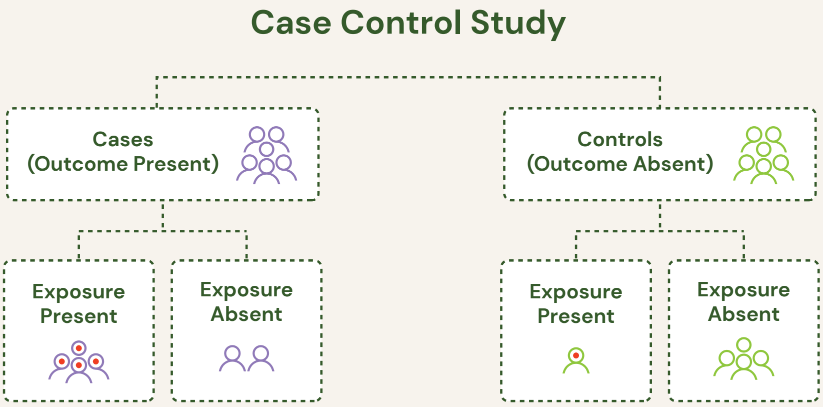
The study did not attempt to change the variables, or to conduct an intervention or to examine cause and effect.

The study shows a snapshot in time of the outcome and exposure; the prevalence of major cardiac defects amongst those infants exposed to Zika virus in utero.

Case control study

A type of observational analytic study. Enrolment into the study is based on presence ("case") or absence ("control") of disease. Characteristics such as previous exposure or risk factors are then compared between cases and controls to determine if there is an association between exposure and outcome.

If you wanted to investigate a foodborne outbreak at a wedding, you would select all of your cases (people at the wedding who had the outcome, e.g., Diarrhoea and vomiting) and all of your controls (people at the wedding who did not have the outcome) and investigate all of the different foods that each group of people had consumed during the wedding. In doing so, you can find an association between particular food exposure/s and the outcome.


Case Control Study

Cases (Outcome Present)

Controls (Outcome Absent)

Exposure Present

Exposure Absent

Exposure Present

Exposure Absent

Advantages and disadvantages of case control studies include:

Advantages	Disadvantages
• Good for exploring rare outcomes • Relatively quick and simple to implement • Multiple exposures can be examined	• Recall bias: people may forget details about their exposures • Difficult to validate information • May be difficult to find an appropriate comparison group (controls)

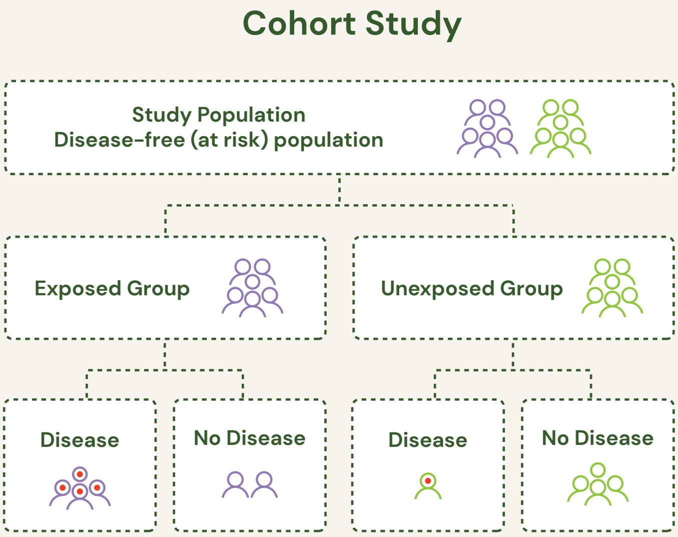
Cohort study

A type of observational analytic study. Enrolment into the study is based on exposure characteristics or membership in a group. Disease, death, or other health-related outcomes are then ascertained and compared.

A cohort study can be prospective or retrospective.

A prospective cohort study is a research study that follows groups of people over time. The groups have many similar characteristics (e.g., sex) but will be different in particular characteristics, such as their exposure (e.g., whether they smoke or not). In a prospective cohort study, the researchers follow the two groups over time to compare who develops a particular outcome (e.g., lung cancer). A prospective cohort study may recruit participants when they are in their 20s, for example, and follow them throughout their whole lifetime to compare if the outcome (e.g., lung cancer) is different between the two groups (e.g., Smokers and non-smokers).

A retrospective cohort study is a research study that compares medical records of groups of people who share some characteristics (e.g., sex) but have other different characteristics (e.g., whether they smoke or not). These groups are compared to see if they have developed a particular outcome (e.g., lung cancer) which has either already occurred or not. A retrospective cohort study would select medical records from people who already have the outcome (e.g., lung cancer) and those who do not and go back through the records to determine if the exposure (e.g., Smoking) is different between the two groups

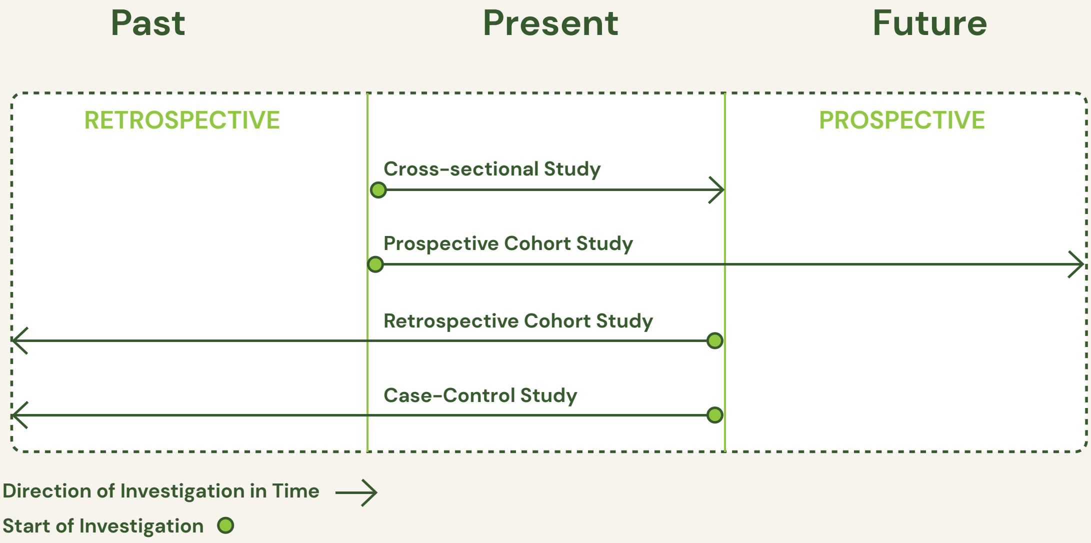


.

Advantages and disadvantages of cohort studies include:

Advantages	Disadvantages
• Can assess causality • Good for investigating rare exposures • Can assess multiple outcomes for an exposure	• Prospective cohort studies can take a long time, be very expensive, and difficult to continue follow-up over long periods of time • Retrospective cohort studies may be influenced by recall bias

The below diagram shows the direction in which the investigation goes for each study type.



Sample

When researching a group of people, it is usually not possible to collect data from every single person in that group. Instead, it is possible to select a smaller group of people who can participate in the study. This is called a sample.

A sample is a selected subset of a population used to measure something about the whole population. A sample may be selected using random or non-random methods, and the sample itself may be representative or non-representative of the larger population.

> **Example:** You are conducting a survey in a village with 1000 people. You only have one day to do the survey, so it is not possible to ask all 1000 people. Surveying 100 people is a sample of your overall population, but not necessarily representative of everyone's responses.

A representative sample has characteristics that correspond to those of the original population or reference population. These characteristics may be related to age, sex, location, occupation, presence or absence of a risk factor or many other possible options.

> **Example:** You are interested in interviewing healthcare workers at a hospital. You know that there are 100 healthcare workers, and 70 of them are women. You may only have time and resources to interview ten healthcare workers. To make this sample more representative of your overall population (all healthcare workers at the hospital) you could invite seven women and three men to be interviewed.

Population
Sample

Image adapted from: https://www.scribbr.com/methodology/sampling-methods/

Sampling methods

There are multiple sampling methods (ways to select a sample). Below is a description of a few of the most common sampling methods used by field epidemiologists.

Random sampling

A sampling method whereby each individual has the same probability of being selected.

> **Example:** You are conducting a survey in a village with 1000 people. You cannot survey everyone from the village so decide to select 100 people as a sample. You want the survey responses from the sampled population to accurately reflect the perspectives of the whole community as much as possible. You decide to take a random sample of 100 people from the 1000 total people in the village. You could enter all the names into Excel and then randomly choose 100 people, for example, or you could choose every second house as you walk around the village. In this way, each person has the same chance of being selected to participate in the survey.

Simple Random Sample

Image adapted from https://www.scribbr.com/methodology/sampling-methods/

Convenience sampling

A sampling method where the researcher selects people who are easily accessible. It is a simple and cheap way to collect data. The data are not necessarily representative of the population.

> **Example:** You want to understand challenges nurses face in clinical work during COVID-19 by interviewing 20 nurses. Whilst it would be ideal to explore the different challenges faced by nurses in different settings (e.g., Rural or urban, private or public facility), you have a short deadline and limited funding which means you cannot travel or pay for participants' travel. Convenience sampling in this example would be to select 20 nurses that work at the health facility you have best access to and would be available for interview in the same city. The sample will not be representative of the whole nursing profession but will give insights into some of the challenges faced by nurses working during COVID-19.

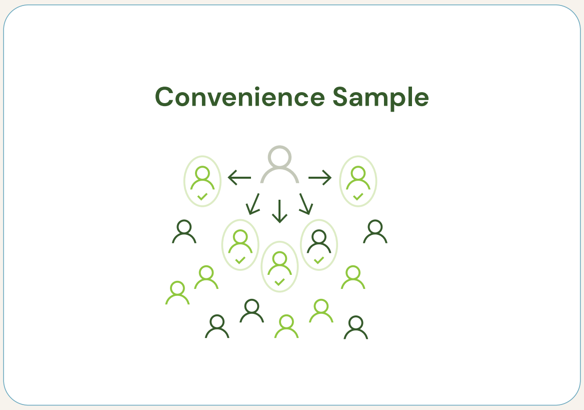


Image adapted from: **https://www.scribbr.com/methodology/sampling-methods/**

Purposive sampling

A sampling method where the researcher selects their sample based on participant characteristics that best meet the purposes of the research. There must be clear criteria for inclusion. It is used commonly in qualitative research studies where the researcher wants to understand the experience of a particular phenomenon, or when the study population is very small (e.g., people with an extremely rare disease).

> **Example:** You want to understand why some patients with tuberculosis become lost-to-follow-up (LTFU). You purposefully select a number of LTFU patients to collect data on a range of experiences.

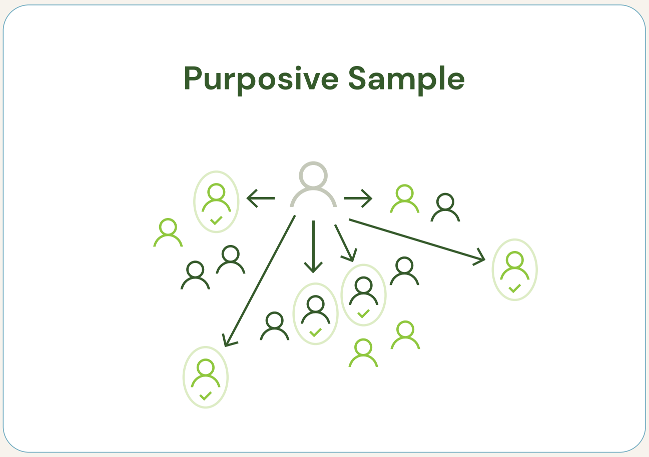


Image adapted from: https://www.scribbr.com/methodology/sampling-methods/

Snowball sampling

A sampling method where the researcher recruits participants via other (already recruited) participants. Snowball sampling is particularly useful if the population is hard to access because the characteristic of interest is rare, there is stigma related to the characteristic, or because there is no organised way of identifying people with the characteristic.

> **Example:** You want to understand how homeless women access reproductive health services in your city. Once you identify one woman who agrees to participate, she informs you of other women in the same situation and how to contact them.

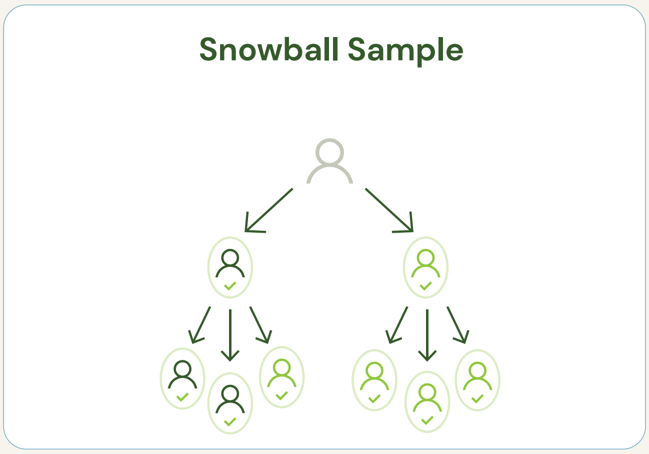


Image adapted from: https://www.scribbr.com/methodology/sampling-methods/

Bias

Bias is any trend in the way that we collect, analyse, and interpret data that leads to conclusions that are systematically different from the truth. Bias limits the conclusions that we can draw from the results of a study.

It is very common for studies to have a degree of error; error can be random or systematic. Random error occurs due to chance and should not affect your study results too much. Systematic error occurs when the same type of error repeatedly occurs leading to bias and potentially an incorrect study conclusion.

Epidemiological studies invariably are subject to at least some bias. Bias needs to be recognised and addressed in study design, planning, and implementation as it is difficult to fix or control bias once data have been collected.

The main types of bias in field epidemiology are **selection bias** (occurs during design and implementation phases) and **information bias** (occurs during data collection). There are many different subtypes of bias; this section will focus on some of the primary types that field epidemiologists need to consider in operational work and project design and implementation.

Selection bias

Selection bias is a result of any error in the selection of participants or factors that may affect participation. The result is that there is a meaningful difference in the relationship between exposure and outcome between those selected to participate and those who do not participate leading to the sample not being representative of the population it is drawn from.

The following types of selection bias are common in field epidemiology and should be considered in study design and implementation:

Sampling bias

> Sampling bias occurs when the study participants are selected in a non-random way. This means that some members of a group are systematically more likely to be selected to participate than others. Sub-populations may be excluded from selection, which can lead to meaningful difference between participants and non-participants. This means that results are not able to be generalised to the broader population; they are applicable only to people with the same characteristics as the selected sample.
>
> To address sampling bias, we ideally select study participants randomly. In this way, each person in the population has an equal opportunity of being selected to participate, and we reduce sampling bias.
>
> In some circumstances (e.g., resource or time limited situations or when the population is small), we may choose a non-random sampling process, such as convenience sampling, for a project. Convenience sampling increases the likelihood of sampling bias because we are intentionally selecting people based on a known characteristic that may be meaningfully different to non-participants. Convenience sampling is commonly used in epidemiological studies; it is important to understand and be able to describe the limitations of this sampling strategy in terms of how convenience sampling may impact the generalisability of the findings.

Non-response bias

> Non-response bias is where the people who do not respond to study invitation or who drop out are systematically different from those who agree to participate.
>
> **Example:** In a project wanting to understand the outbreak investigation experiences of FETP graduates, people living in remote areas with limited or no internet network may be systematically more likely to not respond to project invitations sent by email than those based in urban areas with good internet network. In this case, the outbreak investigation experiences may be very different between urban and remote FETP graduates, and as such, the findings will not reflect the true range of experiences of graduates.

Volunteer bias

> Volunteer bias is when those who volunteer to participate are systematically different to those people who do not volunteer to participate. Participants are more likely to participate in studies and projects where they are interested in the topic being studied. As such, they may not reflect the population as a whole.
>
> **Example:** In a project aiming to evaluate the impact of a community exercise program, people who are already committed to exercising regularly may be more likely to participate than those who do not already have regular exercise practice. This difference may influence the results and our conclusions as those who volunteer for the study are meaningfully different to those who do not.

Information bias

Information bias occurs when exposure or outcome data are collected in a systematically different way between study groups. This can lead to an incorrect estimate of the association between exposure and outcome. Some common types of information bias are explained below.

Recall bias

> Recall bias is a type of information bias. It occurs when one study group tends to recall information differently to the other study group. Information provided by one group may be less accurate than that of the other group. This is a common issue in retrospective studies where people are self-reporting.
>
> **Example:** A case control study where cases are selected based on the presence of the outcome (e.g., a disease). Cases may have thought extensively about their disease, risk factors, and behaviours prior to disease onset. Cases may therefore be more likely to recall “exposures” than the controls who have not had the disease/outcome.

Social desirability bias

> Social desirability bias is a type of information bias. It occurs when respondents answer a survey or questionnaire in a way they think will be seen as better by others.
>
> **Example:** In a study exploring the association between exercise (exposure) and a health event (outcome), participants are asked about their level of exercise. If exercise is viewed favourably in their community they may say they exercise three times per week when it is really only one.

Interviewer bias

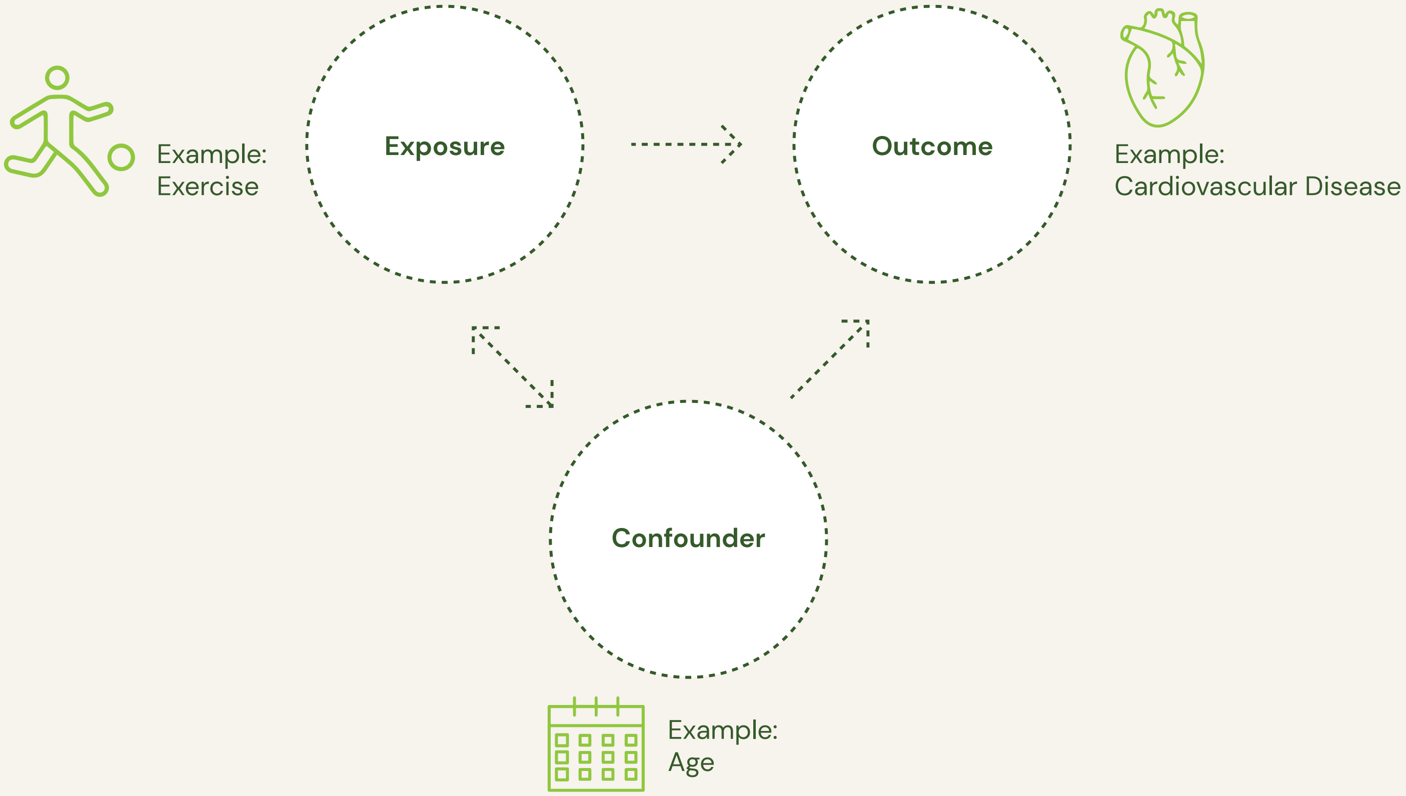
> Interviewer bias is a type of information bias. It occurs when the person conducting interviews in a survey or study influences the responses given by respondents. This can be an issue if questionnaires are not standardised or when the interviewer knows the outcome or exposure status of an individual, causing them to ask leading questions or change the way the questions are asked.

Confounding

Confounding occurs when a third factor provides an alternative explanation for the association between the disease (or outcome) and an exposure (or independent variable).

To be considered a confounder, this third factor needs to be associated with both the exposure in question and the disease (outcome). Confounding can be minimised in the design and analysis phases of an investigation.

> **Example:** Epidemiologists want to study the association between exercise and cardiovascular disease. Age could be a potential confounder. Age is a known risk factor for cardiovascular disease. Also, a person's age could influence their ability to participate in physical exercise. It is important the researchers account for age in their analysis - either analysing the data by age group (stratifying) or using statistical methods to adjust for age. If they don't do this, they might find misleading results which have been confounded by age.



Monitoring, Evaluation and Learning

Monitoring

Monitoring is the routine, systematic monitoring of project resources (or 'inputs'), activities, and results. It is an analysis of these factors to help guide project implementation. Monitoring is concerned with determining if activities are being carried out as planned (i.e., monitoring the process).

Monitoring involves data collection to address questions such as:

- To what extent are planned activities delivered as planned?
- What services are provided, to whom, when, how often, for how long, and in what context?

Monitoring data are used to inform project managers if and where existing efforts need to be modified.

Evaluation

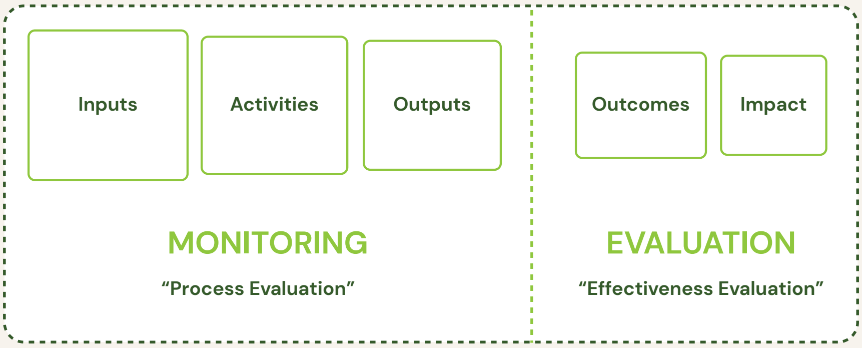
Evaluation is the periodic (including interim and final) assessment and analysis of an ongoing or completed project. Evaluation is focused on the outputs and impacts (i.e., the effectiveness of the project or study under evaluation)

Learning

Learning is the process through which information generated from monitoring and evaluation is reflected upon and intentionally used to continuously improve a project's ability to achieve results.

Item	Monitoring	Evaluation
Frequency	Regular, ongoing	Episodic – at a point in time
Main Action	Keeping track/oversight	Assessment
Basic Purpose	Improving efficiency, adjusting work plan	Improve effectiveness, impact, future programming
Focus	Inputs/outputs, activities/process, work plans	Effectiveness, relevance, efficiency, impact, sustainability
Information Sources	Routine systems, field visits, stakeholder meetings, output reports, rapid assessments	Same as monitoring, **plus** surveys (pre-post-project), special studies
Undertaken By	Project/program managers, community workers, supervisors, community (beneficiaries), funders, other stakeholders	Project/program managers, external evaluators, community (beneficiaries), supervisors, funders
Learning	Used to continuously adjust workplan, make improvements as the project goes.	Highlights effectiveness; used to inform future programming based on what does/doesn't work.
	Used to identify key recommendations to share with other stakeholders doing similar work so that we can improve intervention planning, methods, and implementation.	

Monitoring and evaluation pipeline



The monitoring and evaluation pipeline steps are described in the table below:

M&E Step	Examples	Example question to ask
Inputs	• Staff • Money • Supplies • Training • Facilities • Other resources	• Are we on budget? • How many staff are required to implement our activities? What roles will they have? • How many peripheral resources do we need? • Fuel for boats • Paper for printing surveys • Venues for training • How many vaccines are required to complete our target objective?
Activities	• Vaccinating children • Community awareness sessions • Testing febrile cases for malaria • Interviewing pregnant women about attending antenatal clinics	• Did the vaccine clinic happen? On time? • Did the community awareness sessions happen? • Did the health workers know the case definition to test for malaria? • Did the health workers know how to use the survey? • Were the pregnant women happy to participate in the survey?
Outputs	• # vaccinated out of # in target population • # community awareness sessions held • # people attending session out of # in community • # tested for malaria out of # who met case definition • # interviewed out of # women selected to participate in survey	• How many children were vaccinated? • Did we test everyone who met the malaria case definition? • How many of the 120 women selected, were interviewed?
Outcomes	• Trends of disease • Behaviour changes • Attitude changes • Clinical outcomes • Quality of life	• What was vaccine coverage before the intervention? After? • Have more women had supervised deliveries in Nov 2018, compared to Nov 2017? • What is health worker knowledge about malaria testing practice before the training? After the training?
Impact	• Morbidity • Mortality • Disease trends • Economic impact	• Number of measles cases in clinic A declined from an average of 28 cases/month in 2017 to 2 cases/month in 2019 • Maternal death rate decreased in 2019 by 23% when compared to 2014–2018 • New training program reduced vaccine wastage by 62% and saved $250,000 USD

Monitoring and evaluation matrix

Below is an example of a monitoring and evaluation matrix.

Intervention Objective	Activities	Output Indicators	Outcome Indicators	Source of Information
1 - Objective 1	1.1 – Activity 1	1.1.1 - Indicator 1		
		1.1.2- Indicator 2		
	1.2 – Activity 2	1.2.1 - Indicator 1		
		1.2.2 - Indicator 2		
		1.2.3 - Indicator 3		
2 - Objective 2	2.1 - Activity 1	2.1.1 - Indicator 1		
		2.1.2- Indicator 2		
		2.1.3 - Indicator 3		
	2.2 - Activity 2	2.2.1 - Indicator 1		
		2.2.2 - Indicator 2		

Below is an example of a monitoring and evaluation plan matrix where the goal is to reduce malaria incidence on the island.

Intervention Objective	Activities	Output Indicators	Outcome Indicators	Source of Information
To improve malaria prevention habits amongst residents on the island	1. Distribute 200 bed nets to households on the island	1.1 # bed nets distributed (%) 1.2 # (%) houses on the island to receive their first bed net	% bed net coverage compared to % before intervention	• Survey of households visited to distribute nets • Audit of nets distributed
	2. Conduct 4 community education sessions about malaria prevention methods	2.1 # community education sessions conducted 2.2 Total # of people who attended community education sessions (% of estimated # residents of the island)	Participants knew % of basic malaria prevention steps after sessions compared to % before	• Log of community educational sessions, including # of participants • Pre & post assessments of participant knowledge, attitudes & practices
	1 & 2		% positive malaria tests in health facility compared to % before the interventions	• Health facility records

Below is the completed monitoring and evaluation matrix for the above plan, where the goal was to reduce malaria incidence on the island.

Intervention Objective	Activities	Output Indicators	Outcome Indicators	Source of Information
To improve malaria prevention habits amongst residents on the island	1. Distribute 200 bed nets to households on the island	1.1 **124** bed nets distributed **(62%)** 1.2 **35 (28%)** houses on the island to receive their first bed net	**94%** bed net coverage compared to **84%** before intervention	• Survey of households visited to distribute nets • Audit of nets distributed
	2. Conduct 4 community education sessions about malaria prevention methods	2.1 **2** community education sessions conducted 2.2 **96** of people who attended community education sessions (**8%** of estimated **1200** residents of the island)	Participants knew **96%** of basic malaria prevention steps after sessions compared to **90%** before	• Log of community educational sessions, including # of participants • Pre & post assessments of participant knowledge, attitudes & practices
	1 & 2		**45%** positive malaria tests in health facility compared to **42%** before the interventions	• Health facility records

Field Epidemiology
IN ACTION