# Principles of Epidemiology

**A textbook on the principles of epidemiology.**

By H. Giovanni Antunez
St Cloud State University

# Contents

PART I

# PART I

Introduction and Bio

# Author's Biography

## H. Giovanni Antunez

H. Giovanni Antunez earned a MD in general practice at the National University of Honduras, Medical School, a MPH and a DrPH from Tulane University School of Public Health and Tropical Medicine, New Orleans, and it is a also a Certified Health Education Specialist (CHES). He has been teaching epidemiology (and other public health courses) for 18 years at St. Cloud State University, where he works at the Community Health program, Dept. of Kinesiology, College of Wellness and Health Professions. His background on teaching public health courses, including epidemiology came from the mix of degrees that he had obtained, and from the lots of experiences teaching public health courses online and face to face. He recognized that a textbook in epidemiology for undergraduate students is a great need. This textbook is a response to this need. He teaches courses on the area of epidemiology, public health, nutrition, health promotion, environmental health, international health, consumer health, and others. He also provides academic advising to community health students and is engaged in several campus wide activities besides teaching.

# Introduction

**Welcome!**

Welcome to this book which has been written with you (the student) in mind. I hope this becomes a useful tool for you to get a grasp to understand the field of epidemiology.

'Open book', image from Wikimedia Open Clip Art Library, Public Domain.

**For whom is this book for?** Most books about epidemiology are written for graduate students but this book is especially designed for undergraduates, who are taking an introductory course in epidemiology, and especially for those students who want to know more about epidemiology. But not because it is directed mainly to undergraduates and not to graduate students the course has less rigour and quality, it is because the concepts are introduced, and presented in an easier and simplified form, so, the study of epidemiology becomes less intimidating. While teaching epidemiology for several years, I have found that some students don't choose epidemiology as a career because the content seems too complicated, presenting the content in a practical manner, and simpler (although some epidemiological concepts are very complex); should motivate students to learn about this subject matter.

**Why did I feel the need to write this book?**

This book is written in response to the need to free open access resources that allow students to learn more about a subject area, in this case, epidemiology. A textbook is a must for an epidemiology class but I have several students who don't have the financial resources to buy a textbook. It is accepted that having a textbook to do the readings is a tremendous help for a successful outcome in a college course. Personally, I cannot imagine how to study epidemiology without having a textbook, and only relying in class notes, instructor's power points, and some pieces of epidemiology found in the Internet. Those are the main reasons I am writing this book.

*Main purpose of this textbook*

The main purpose of this textbook is to introduce students to the field of epidemiology, especially for those student who may choose epidemiology as a career.

**It is said that knowledge is power**

The knowledge of epidemiology can be used to better understand medical and public health journal articles, especially the data and results section, which usually include numerical information layout on tables that contain statistical results, a basic knowledge in epidemiology can help with this.


Image "Knowledge" by Nick Youngson CC BY-SA 3.0 Alpha Stock Images

A knowledge in epidemiology can also help to understand the concept of rates, ratios and counts and also prevalence, incidence, odds ratio, and relative risk that are the main concepts learned in an introductory course in the field. Once these concepts are understood, the data start making sense and the overall information contained in medical and public health journals can be used more efficiently.

**What I think overall**,
I think that taking a basic course in epidemiology is a tremendous tool to make better sense of the information that is published in scientific journals, and even in a lot of popular readings and publications found in the World Wide Web. It is my hope that this small book will help you (the student) to becoming interested, and maybe passionate about the field of epidemiology and its role in medical and public health practice.

**Acknowledgments**
I want to acknowledge the support of the Minnesota State Higher Education system to allow me to participate in the Open Education Resources (OER) Initiative that intents to provide free access textbooks to its students in these current times in which textbooks have become so expensive.

Thanks also to Dr. Karen Pikula, who have been my mentor and facilitator at the Minnesota State OER Learning Circles in which I have participated in recent years, she believed in me that I was able to prepare this content and complete the work.

Also, thanks to the administrators in St Cloud University who has been supportive signing the OER Learning circles paperwork, and welcoming of the OER textbook's Minnesota State initiative, which intents to close the equity gap that will make higher education available to all students in the system. To all, thanks,
Giovanni A.

PART II

# PART II

# 1. What is Epidemiology?

***Learning Objectives***

By the end of this chapter, the learner will be able to

- Define and understand the field of epidemiology
- List the areas of study in epidemiology
- Understand epidemiology as an interdisciplinary science
- List the main uses of epidemiology in public health and social sciences
- Explain how the field of epidemiology started and its contributions to the study in causality and disease

### What is epidemiology?

The field of epidemiology is relatively a new science, at the point that sometimes (probably frequently), it is better understood in the context of another field of study, biostatistics. But epidemiology has its own terminology, and also its own methodology that make it a little different, if not very different from biostatistics.[1]

### Epidemiology Definition

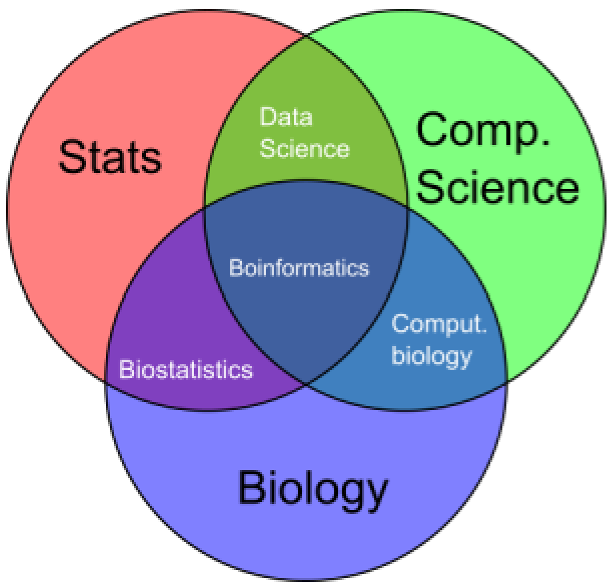
It is accepted that , epidemiology is *the science to understand, and explain the occurrences of health phenomena in a population.* As a branch of public health, epidemiology studies not only disease, but also different health outcomes. [2]

### Epidemiology: an interdisciplinary science

1. No Author. (n.d.). Epidemiology vs. Biostatistics: Understanding Their Roles in Public Health. University of Nevada Reno. Available at: https://onlinedegrees.unr.edu/blog/epidemiology-vs-biostatistics/
2. Hernandez JBR, Kim PY. Epidemiology Morbidity And Mortality. (2020 Oct 13). In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan–. PMID: 31613448. Available https://www.ncbi.nlm.nih.gov/books/NBK547668/

Since epidemiology integrates several aspects of the social sciences, and at the same time uses information, and tools from other sciences, that is the reason to consider, *epidemiology as an **interdisciplinary** science.* It takes information from statistics, medicine, biology, sociology, nutrition, cultural studies, gender studies, anthropology, etc., and integrates them in the production of models, formulas, and analyses that can be used to predict, prevent, and analyze many health phenomena, including the issue of quality of life, and wellness. Its contributes are vast, because of its interdisciplinary nature.[3] At the same time for epidemiology to be interdisciplinary brings some complications, since there is no way that results are free of errors; but as a science that auto-correct through trial and error, make it stronger for advancement.[4]

See an image (representation) of how interdisciplinary sciences work for a degree in bioinformatics.




A venn diagram showing what you might find at the intersections of statistics, computer science and biology. Image from [Genmejigsaw](), [Bioinformatics]().

3. Hanson, A., & Levin, B. (2018-02). Epidemiology, Chapter 4. In Introduction to Public Health in Pharmacy. Oxford, UK: Oxford University Press. Available at: https://oxfordmedicine.com/view/10.1093/med/9780190238308.001.0001/med-9780190238308
4. Schoenbach, VJ. (2000). Chapter 4. Phenomenon of disease In Epidemiology Net. Available at: http://www.epidemiolog.net/evolving/PhenomenonofDisease.pdf

### Common Uses of Epidemiology

The following are commonly accepted uses of the field of epidemiology, details of each are including in the table below:

| Use(s) | Example(s) | Additional examples and comments |
|---|---|---|
| To predict and prevent adverse (or, positive) health outcomes. | Epidemiology can be used to predict, and prevent adverse outcomes especially with those situations that are complex health issues. | The prediction feature has been one of the main characteristics because it is used to forecast for example, **epidemics**; and the prevention part confirms that epidemiology is an essential part of public health. |
| To determine the spread, and distribution of health issues | Epidemiology can be also used to determine the spread, and distribution of health issues, and outcomes in a community, by providing not only the data that supports the report, but also, the baseline information that policy makers can use to support current public health policies, and in some cases laws that regulate the unhealthy behavior. | In other cases, it can be used to promote a healthy behavior in order to control, or, prevent the disease (as it has been mentioned before in this section), or, the health outcome (not necessarily a disease) that contributes to the overall quality of life of the affected group, or, population. |
| To collect information, and track different health outcomes. | Epidemiology is used also to collect information, and track different health outcomes in a population. The most common method to conduct this function, is known as **surveillance** system that is used to systematically collect, analyze and disseminate health data that is commonly used for the functioning of public health programs, and services. | More than a method it can be used also to allocate resources that will help the public health practitioner find the needed resources for specific health conditions. |
| To study infectious diseases and also chronic diseases. | Since epidemiology started with the study of infectious diseases, the field was used mainly to study them. | the field of epidemiology expanded, and it also included the study of chronic diseases such as cardiovascular and cancer, which are also called, **life-style diseases** since their risk factors are linked to the way in which people live. |

| Use(s) | Example(s) | Additional examples and comments |
|---|---|---|
| The planning of health services In this case, epidemiology is used also for the planning of health services, not only in clinical settings; but also at the population level, or, public health level (programs). | For example, an epidemiological study about vaccination in an low income, and low literacy community will provide, and generate information, and data that can inform immunization public health programs. | Also, it can also inform (especially data) policy makers who want to legislate for more fair access to vaccines, especially to those groups in the population that have low, or, no access to these services. |
| Epidemiology and the evaluation of public health programs. | Epidemiology is very useful for the evaluation of health programs, including community-based programs, and in general health initiatives (such as coalitions, and consortia), in which an evaluation of the impact of the interventions is needed in order to keep the program alive, by getting funded. | Most funding organization these days, require an evaluation of the impact of most program's intervention, and epidemiology can assist with that in many ways by providing the numbers (statistics) that **quantify** the overall impact, but also the data that can be used also to predict health outcome. |
| For Epidemiology fieldwork. | In recent years the field of epidemiology has been opened to include **qualitative** data (and not only numbers, and it did in the past) to enhance, and put a face to the numbers; the most common of these are case studies, in which the number is small, and the meaningfulness of the information is accomplished by using qualitative data such as In depth interviews, personal stories, and narratives. | As we had mentioned in other sections of this chapter and the rest of the textbook, epidemiology by nature has been a quantitative science. |
| For the design and implementation of data collection instruments | In addition epidemiology can assist the public health, and medical practitioner with the development of data collection instruments such as surveys, and questionnaires commonly used to obtain population data that will be later analyzed using the epidemiological approach, which is population based. | Most of these data collection is quantitative. |

| Use(s) | Example(s) | Additional examples and comments |
|---|---|---|
| To address public health problems. | Health outcomes are linked to poverty, low socioeconomic status, access to resources, issues of racism, and discrimination, etc. These are important aspects of public health. | As part of public health, epidemiology has an important role in social justice too, because many of the adverse health outcomes are linked to social issues such as poverty, low socioeconomic status, access to resources, issues of racism, and discrimination, etc. |

The content of this table has been prepared using the references listed here. [5], [6], [7], [8], [9], [10], [11], [12], [13], [14].

The list of uses of epidemiology included in this book does not pretend to be exhaustive. There are probably more uses that has not been listed here but they can be found in the medical and public health literature.

5. Frérot, M., Lefebvre, A., Aho, S., Callier, P., Astruc, K., & Aho Glélé, L. S. (2018). What is epidemiology? Changing definitions of epidemiology 1978-2017. PloS one, 13(12), e0208442. From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6287859/
6. Centers for Disease Control and Prevention. (CDC). (2021). Social Determinants of Health: Know What Affects Health. Available at: https://www.cdc.gov/socialdeterminants/index.htm
7. Thacker, S. B., Parrish, R. G., & Trowbridge, F. L. (1988). A method for evaluating systems of epidemiological surveillance. World health statistics quarterly. Rapport trimestriel de statistiques sanitaires mondiales, 41(1), 11–18. From https://pubmed.ncbi.nlm.nih.gov/3269210/
8. Al-Maskari, F. (n.d.). Lifestyle diseases: An Economic Burden on the Health Services. United Nations, Chronicle. From https://www.un.org/en/chronicle/article/lifestyle-diseases-economic-burden-health-services
9. Centers for Disease Control and Prevention. (CDC). (2011). Lesson 1: Introduction to Epidemiology Section 1: Definition of Epidemiology In Principles of Epidemiology in Public Health Practice. Third Edition. Available at: https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section1.html
10. Centers for Disease Control and Prevention. (CDC). (2011). Epidemiology Glossary. Available at: https://www.cdc.gov/reproductivehealth/data_stats/glossary.html
11. Wolff, B., Mahoney, F., Lohiniva, AL., Corkum, M. (n.d.). Collecting and Analyzing Qualitative Data In the CDC field epidemiology manual. From https://www.cdc.gov/eis/field-epi-manual/chapters/Qualitative-Data.html
12. Goldberg M, Melchior M, Leclerc A, Lert F. (2003 Sept). Epidémiologie et déterminants sociaux des inégalités de santé [Epidemiology and social determinants of health inequalities]. Rev Epidemiol Sante Publique, 51(4):381-401. Available at: https://pubmed.ncbi.nlm.nih.gov/13679732/
13. Kesteloot H. (2004). Epidemiology: past, present and future. Verh K Acad Geneeskd Belg. 66(5-6):384-405. From https://pubmed.ncbi.nlm.nih.gov/15641567/
14. Al-Maskari, F. (n.d.). Lifestyle diseases: An Economic Burden on the Health Services. United Nations, Chronicle. From https://www.un.org/en/chronicle/article/lifestyle-diseases-economic-burden-health-services

**Key Takeaways – Epidemiology is used to**

- predict and prevent adverse (or, positive) health outcomes
- *collect information, and track different health outcomes*
- To study infectious diseases and also chronic diseases
- For the planning of health services
- And other that had been listed above in this section.

### Epidemiology and its beginnings

In addition to all of the mentioned uses and applications of epidemiology since its beginnings have made great contributions to the study of disease causality in medicine and public health. The initial **models** of causality. The most common models are discussed in this chapter. And, the applications of most of the terms will be covered in a brief chapter dedicated to infectious diseases later in the book.

### Approaches used to explain causality

The history of epidemiology is linked to the causality model commonly used in the study of infectious diseases. And, one of the initial approaches used to explain causality is the interaction between host, agent and environment.

### The epidemiological triangle

In many ways, epidemiology started with the famous epidemiological triangle, which is an old model to explain causality. In this concept, for disease to occur, there is need to have a *host*, an *agent*, and an *environment*.[15] This explanation holds true for a long time in sciences, and at a certain point, the variable, *time* was added, in order to add a new dimension, which means, host, agent and environment occur in a specific time that can be minutes, hours, days, etc.

15. CDC. (n.d.). Lesson 1 Understanding the Epidemiologic Triangle through Infectious Disease, pdf material. From https://www.cdc.gov/healthyschools/bam/teachers/documents/epi_1_triangle.pdf

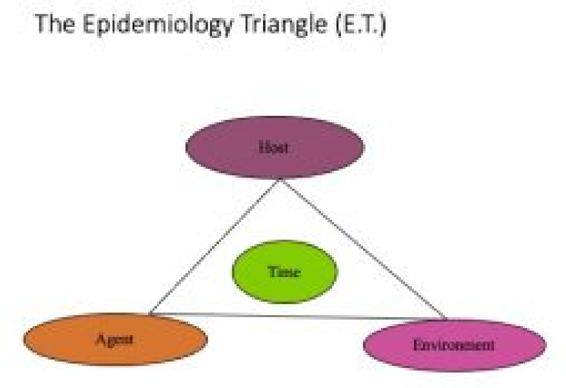



Image prepared by Giovanni Antunez, Licensed [CC BY 4.0]

In so many ways, this model holds true for infectious diseases because it is based on the medical model to explain causality. The situation changes when these concepts are applied to the causality of non infectious diseases, in which the interaction of so many other factors, some of them very unpredictable cannot be explained by this concept of the epidemiological triangle. But for now, in this chapter, we will use the model to introduce the concepts, and definitions commonly used in epidemiology.[16]

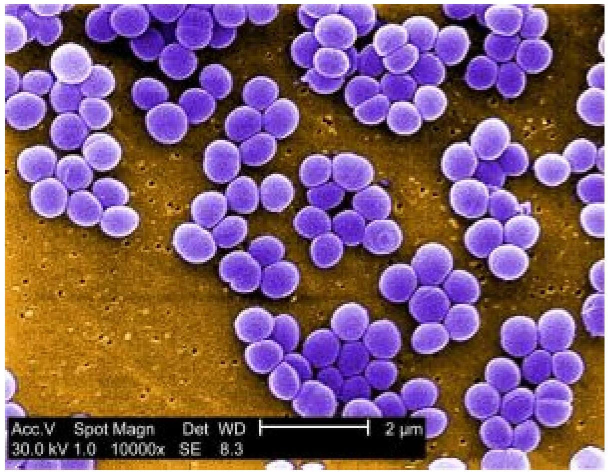
*Key Takeaways (Sidebar)*

As it is clear in this concept of the epidemiological triangle, the conditions for a disease to happen requires an agent, a host, an environment and the time needed for the disease to by manifested.

### The agent

It is accepted that the agent is one of the major causal factors – and of course – in this medical model, common agents include (what we know so far in sciences): bacteria, viruses, fungi, other microbes and parasites. Another common name for the agent, is **pathogen.**pathogen. Examples of common agents, bacteria. See picture below:

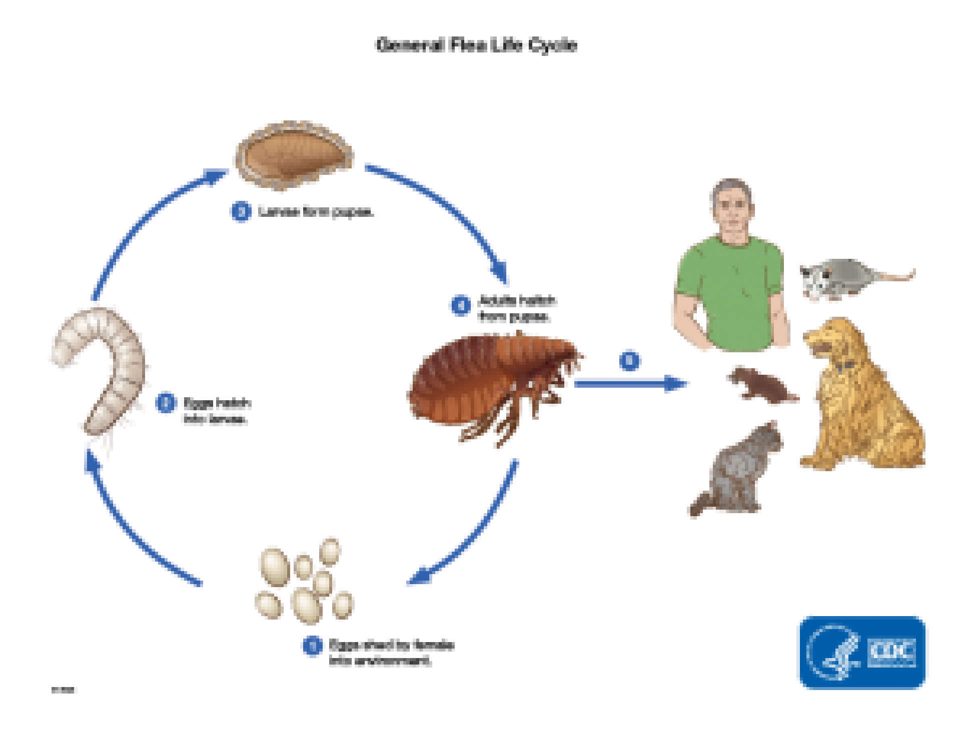
16. Rivier University. (n.d.). What Is the Epidemiologic Triangle? From https://www.rivier.edu/academics/blog-posts/what-is-the-epidemiologic-triangle/




Example of an agent. Staphylococcus aureus bacteria. Image from PIXNIO. Public Domain.

### The host

In the context of the epidemiological triangle, a host can be a person, an animal that harbors the disease. As described later in the textbook, a host can show symptoms of the disease, or, can be free of symptoms.




The host is explained in this diagram shows [‘the typical lifecycle of fleas’](). Image from [CDC](), [Public Domain.]()

### The environment

In general it is accepted that the environment is everything that surrounds the host, or, promote the existence of the agent. It is is the external (to the host or, agent) set of factors that contribute to the development of the disease.

Example of an environment, “Pollution in the city of Los Angeles”. Image from Wikipedia Commons. Licensed CC BY-SA 3.0

### Time

It refers to the duration of the disease, it includes incubation periods (not symptoms are manifested during this period), or, the manifestation of the whole set of symptoms that characterize the disease.

Time, image by Mohamed Hassan, Pixabay

### Other models to explain causation in epidemiology

There is another model commonly used in epidemiology, which is very similar to the epidemiological triangle, and this is called, the ‘Chain of Infection.’ Again, the model works well for infectious diseases but not for other type of **diseases**, or, **disorders**.

### The chain of infection model

In the chain of infection model, the main factors that contribute to disease development are the same, but they seem to be more developed, for example, the agent/pathogen is called, **etiologic** agent/pathogen; there is the addition of another category called, the source/**reservoir, mode** of transmission, and host. See image:

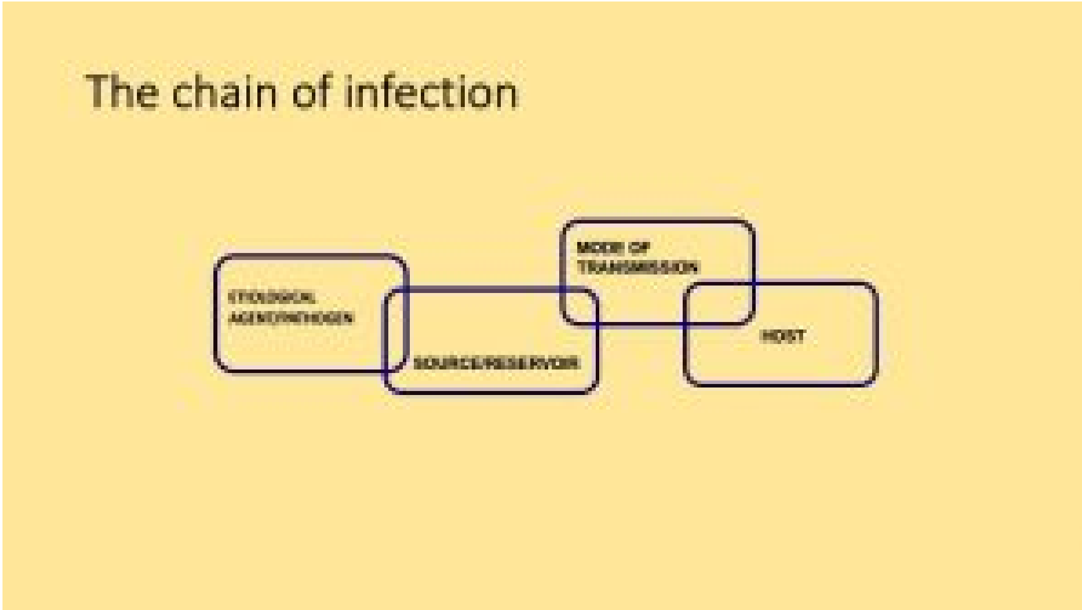



Image prepared by Giovanni Antunez, Licensed CC BY 4.0

**The improved model**

Over the years, the chain of infection model has been expanded and some of the components [although they are essentially the same as the original conceptualization of the model], they appear listed independently, for example, the mode of transmission remains, but two new concepts added, the 'means of **entry**' and the 'means of escape' – which are essentially related to the 'mode' of transmission. Also, the 'host susceptibility' has been added too, and the word, '**organism**' have been added to the the word 'pathogen,' creating a category known as the 'pathogenic organ**pathogenic**ism'. The reservoir category remained unchanged. See image that represent the model:

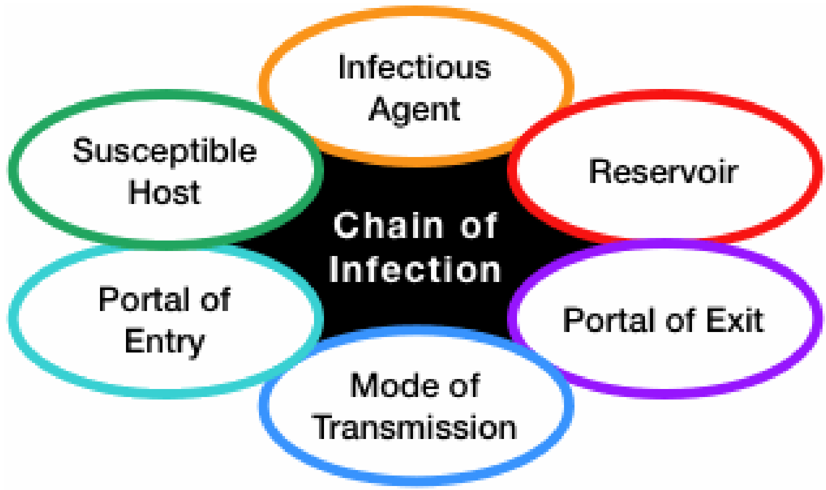



Image from a [Chain of infection]

**Benefits of the two models – epidemiological triangle, and chain of infection**

It is fair to ask, how these models help to understand causality in epidemiology? The answer to this question is that these model complement each other, or, expand on the concepts use to explain how disease transmission occurs. There is a lot details involved in the implementation of this model, but for this section, the major concepts have been discussed. It is also important to note that with the COVID-19 pandemic many of these concepts have re surged because the pandemic is caused by an infectious disease agent.[17]

**The newly developed epidemiological triangle**

In the context of this section, here is how the concepts presented above have contributed to the 'newly improved epidemiological triangle,' which is presented below:

17. Tsui, B., Deng, A., & Pan, S. (2020). COVID-19: Epidemiological Factors During Aerosol-Generating Medical Procedures. Anesthesia and analgesia, 131(3), e175–e178. From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7302069/ or, https://doi.org/10.1213/ANE.0000000000005063

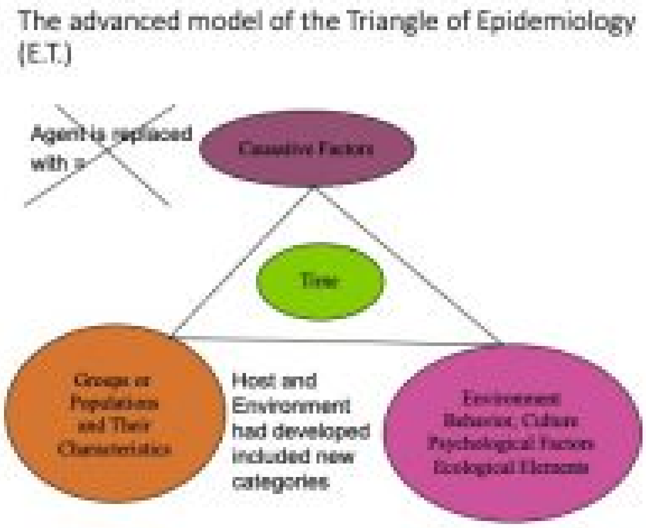



Image prepared by Giovanni Antunez, Licensed [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

The epidemiological triangle is considered foundational to the field of epidemiology, it is limited (very basic) and only applies to infectious and other communicable diseases. For this reason, the need to develop more advanced models have been the basis to develop another concept, which is the web of causation.

**The concept of web of causation**

The main premise of this approach is that there is more than one factor involved in the development of what we call, 'disease.' It is a series of factors that are involved in this process, it is multifactorial. **multifactorial** And, when the several factors are put together, the model looks more like a web -hence the name, web of causation. One example could be the case of use of prostate-specific antigen (PSA) test used to assess if a man need or, not cancer treatment, especially unnecessary cancer treatment.[18] See image below:

18. No author. (n.d.). Example web of causation for PSA test. Wikimedia Commons. From https://commons.wikimedia.org/wiki/File:Example_web_of_causation_for_PSA_test.jpg

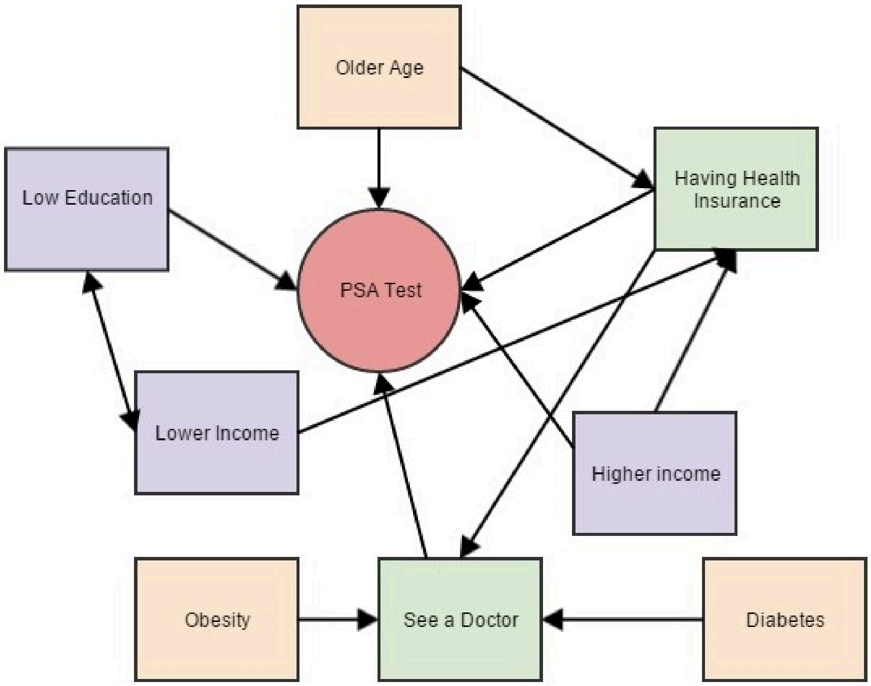



Image 'Example web of causation for PSA test'. Monika Wahi , Wikimedia Commons. Licensed CC BY-SA 4.0

**The concept of Health Determinants**

Although the web of causation (also called, 'web of causality') is very useful, but when there is time to **operationalize** the multitude of factors involved in the process. Most models fall short because it is more than the interaction between an agent, a host, and an environment, and even if the concepts are expanded such as in the case of the chain of infection model. There is need to integrate the concepts of risk factors and other several factors that affect the health of the individual and the community in a direct and indirect form. The concept of health determinants is depicted in the following image:


Image: [Social Determinants of Health](). From [CDC,]() CC license unknown.

In general social determinants are conditions in the places where people live, learn, work, and play that affect a wide range of health and quality-of life-risks and outcomes. [19] This approach is being found useful, especially to understand how health and disease are dynamic concepts that go beyond the identification of agents/pathogens, host characteristics, and the environment. On the other hand, and to be fair, the previous ('older') models can still be seen in this new model, especially that expansion on the concept of environment, and its associated factors.

**Additional Common Epidemiological concepts**

In this section of the chapter, the following concepts will be defined and explained: epidemic, endemic, pandemic, holoendemic, and others.

19. CDC. (n.d.). Social Determinants of Health: Know what affects health. From https://www.cdc.gov/socialdeterminants/

### Epidemic

The term **epidemic** in general refers to an increase in the frequency of a disease (the term applies mostly to infectious diseases, although in recent years, it is common to hear the use of the term for non infectious diseases such as obesity). Or, put in another way, epidemic refers to an 'excess' in the number of cases of a disease. See a general graph (or pattern) of an epidemic:

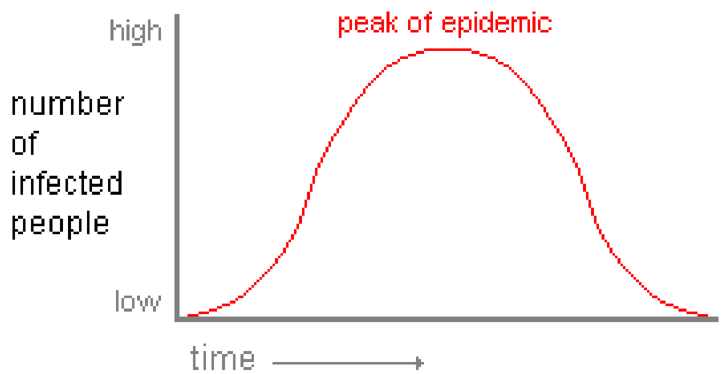



Image from Medical Anthropology epidemiology.

There are several definitions of epidemic found in the Internet, but most of them are confusing, the key concept is the word, 'excess' which also means that there is a baseline in the occurrence of the disease, and once the baseline grows in size, the epidemic has started. The graph shows how the number of cases ascend (or, increase)

2: This graph shows classic logistic growth of SI epidemic model with... | Download Scientific Diagram

Image from Research Gate, originally developed by Mikayel Poghosyan

At the same time, it is common for newly epidemiology students to confuse the term, ‘seasonal disease’ (explained later in this section) with an epidemic disease. But the difference is that a **seasonal disease** occurs during certain seasons, or, times of the year, and then, it declines. A common example is the seasonal flu, which usually starts by early October and ends around April or, May. Then, the number of cases declines significantly (almost disappear) during the summer months usually June – August. See a graphic representation of a seasonal disease such as the flu.

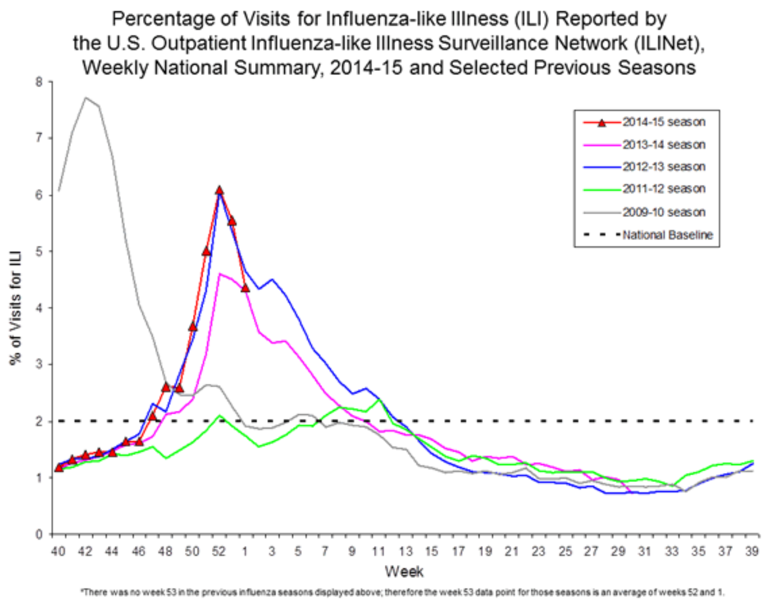
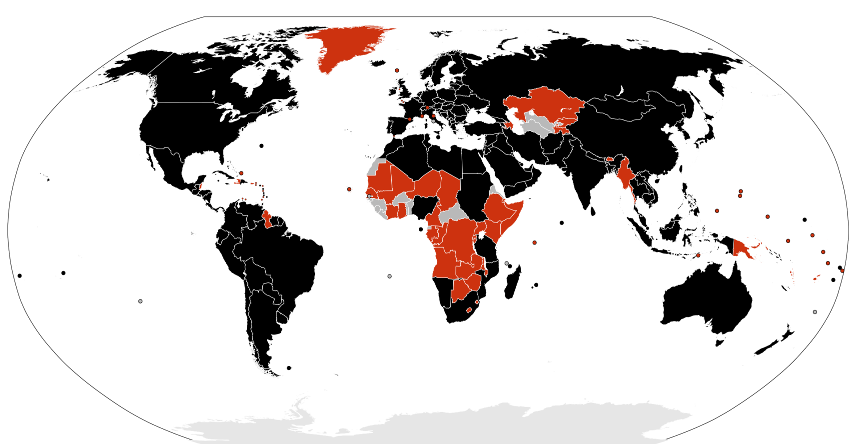


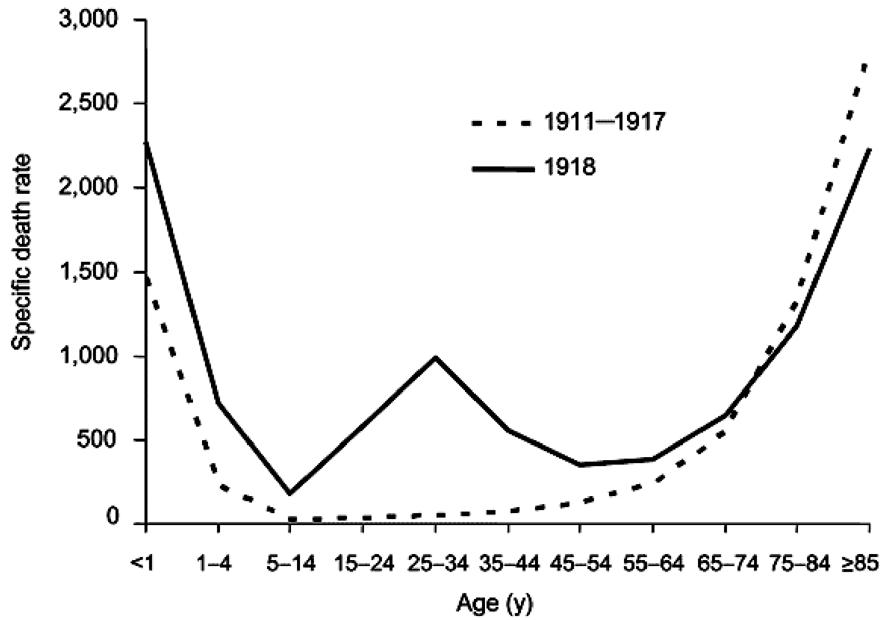

Image from CDC. Influenza outpatient visits, Public Domain.

Another common term, especially in these days of COVID-19, it is **pandemic**. Since a pandemic is a disease that after being epidemic had crossed borders and becomes a health problems across the world, see an image of this below  from the 2009 swine flu pandemic:



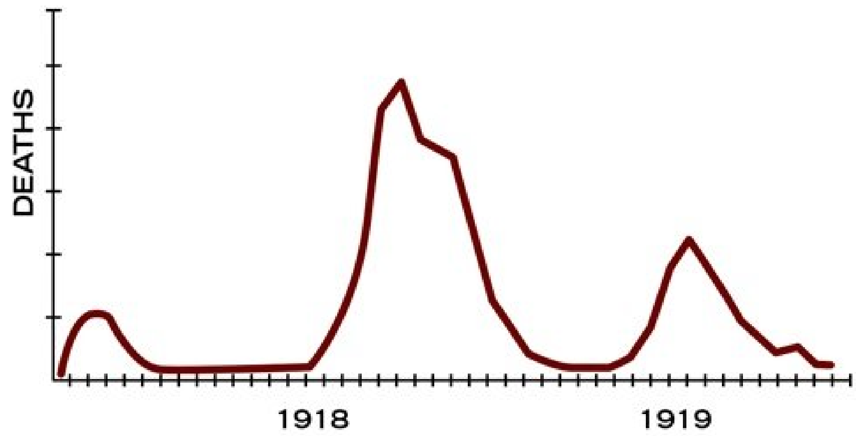
"The 2009 swine flu pandemic by country." Image by 'Allstrak, Wikipedia, licensed CC BY-SA 4.0

There is also a typical (classic) shape of a pandemic, which compared to the graph on an epidemic (see, epidemic here in this chapter), the pandemic comes (goes) in waves, and the entire process passed, it resembles the shape of a 'W' letter. See image below:



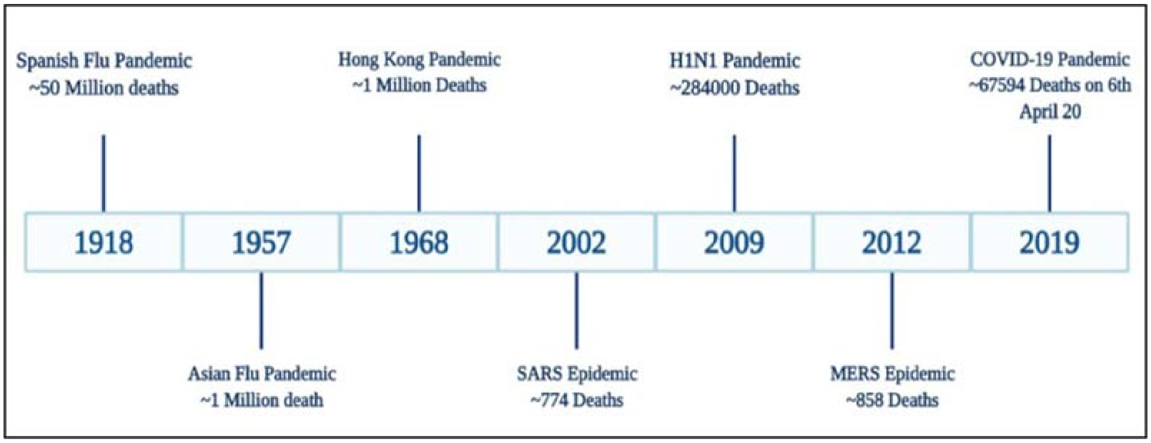

**Spanish Flu by Age**: M.A. Thesis, 2008 Keirsten E. Snover "The Spanish Influenza Epidemic of 1918-1919: The Spokane Experience", Public Domain.

Another image that reinforce the comments about 'pandemic goes in waves,' can be seen below:




"Learning from the Spanish Flu", from Creative Circle Media, Public Domain.

Another characteristic is that pandemic tends to appear every one hundred years, or, closer to one hundred years from the last pandemic as it is the case of the Spanish flu. Epidemiologists use this type of information to compare and predict current and future pandemics. See image below from :




"Timeline of influenza and coronavirus outbreaks since 19th Century." From the article, A Comparative Analysis of the Spanish Flu 1918 and COVID-19 Pandemics In the Open PH Journal. See detail reference.[20] Licensed CC BY 4.0

An additional concept that usually is discussed in this context, it the concept of **endemic**, which essentially refers to a disease with a constant frequency that is limited to a specific area of region. An example of an endemic disease for example in tropical regions, or, countries in the world is malaria, which is 'endemic' for example in some regions of Africa and Latin America, see image below:

20. Agrawal, A., Aadesh, G., Kaivalya, D., Supriya, K., Punit, F., Nazli, K. (2021). A Comparative Analysis of the Spanish Flu 1918 and COVID-19 Pandemics. Open Public Health Journal, 14 (1), 128-134. From https://doi.org/10.2174/1874944502114010128

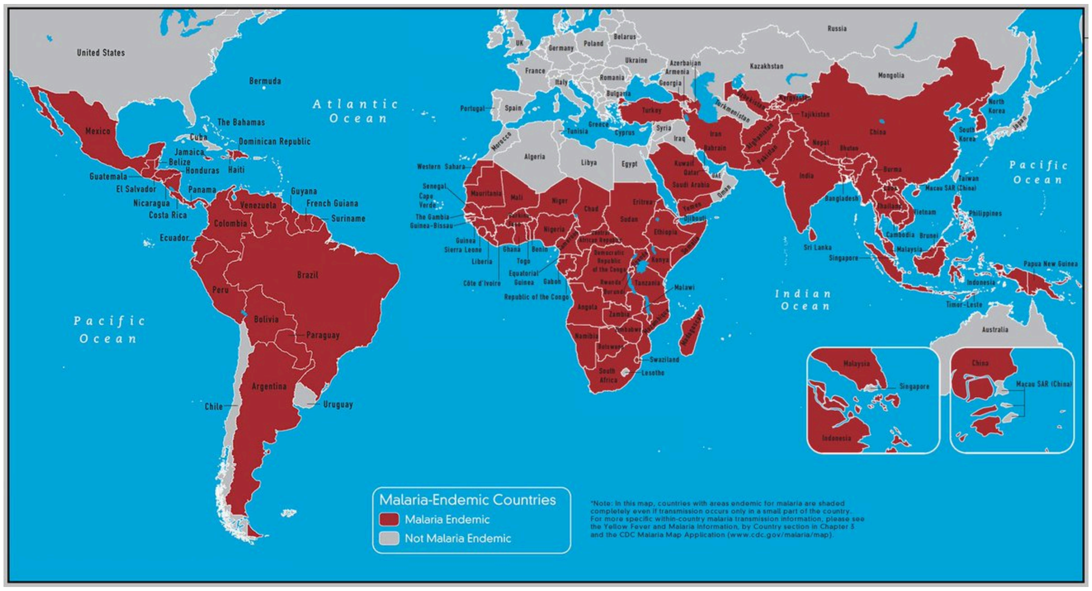



Image found in a [Tweet from Dr. Stavros Saripanidis.]() Labeled CDC map of Malaria Endemic Countries.

Finally, an additional terms is also part of the list of concept discussed here, there are two terms, **hyperendemic** and **holoendemic**, the use of these terms is infrequent but it probably helps to understand that a combination of terms is also possible.

### Other terms and concepts used in epidemiology

The following is a section that covers other concepts, and terms that are commonly used as the foundations of epidemiology, especially for the understanding of infectious diseases. They may not apply to non-infectious diseases since this is part of the original model of epidemiology as it has been mentioned several times in this textbook.

### The natural course of a disease

From its beginning to end, each disease has a natural course (or, 'natural history'). This can be observed if no medical intervention (again, under the medical model for which the field of epidemiology was originally designed) is taken and the disease is allowed to run its full course. As many epidemiologists would agree, the natural history of disease is well-documented for some diseases but not well understood in others. The following image summarizes (and, outlines) the main components of the model:

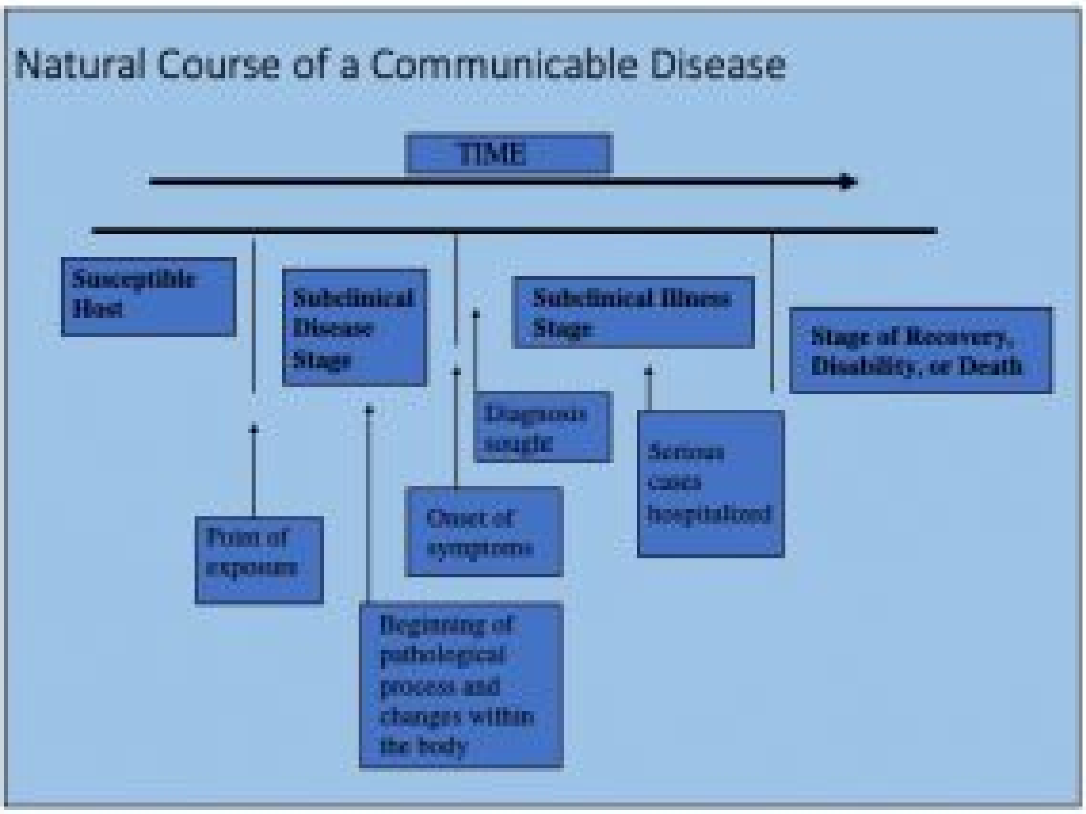



Image prepared by Giovanni Antunez, Licensed [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

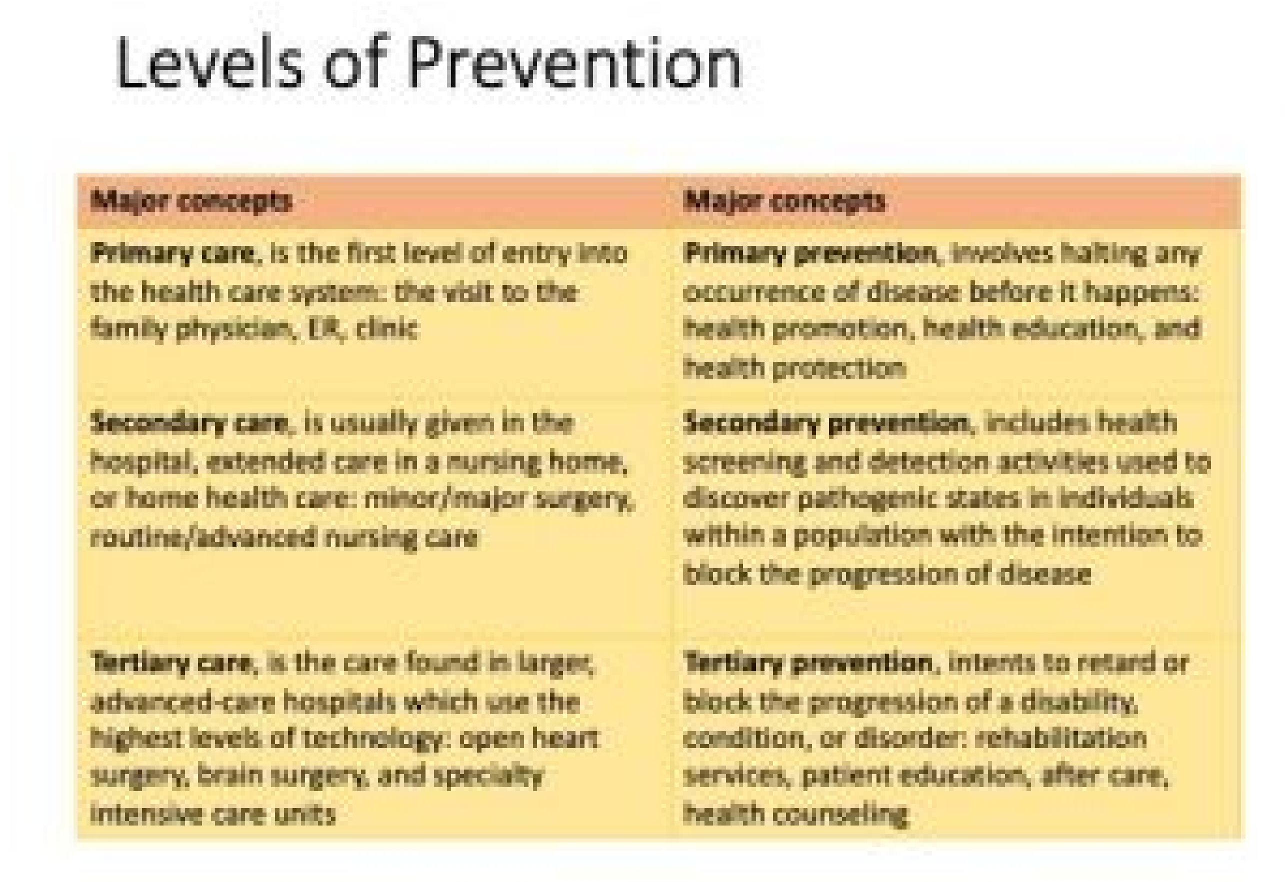
This model (of natural history of disease) has been used **to develop the concepts of primary, secondary and tertiary prevention** that is commonly used in medicine, public health and especially in public health education. And this makes sense if the model again is applied to an infectious, or, communicable disease.

The susceptible host matches the majority of situations known as **primary prevention**,**primary prevention** which is composed by a series of interventions, mainly medical and public health actions that help to prevent diseases, for example, the use of pap smear to prevent and detect early stages of cervical cancer, mammograms, digital rectal exams, colonoscopies, and other common screening tests. The idea is to identify the risk, or, the disease in an early stage mostly before symptoms develop.

When the symptoms start developing, and the person seeks a health care provider, then, the step is referred as **secondary prevention**, this is the most expensive stage because the cases that are serious or, need intense treatment are hospitalized to receive a more intensive intervention, those cases that are not hospitalized are usually given treatment under supervision especially for middle cases of a disease. In the secondary prevention step, the person recovers, dies, or, became disabled, or, with severe limitations such in the case of after a stroke in which most affected individuals have problem with speech, and walking to just mention some of the complications of an after stroke episode.

The last step is called, **tertiary prevention**, which occurs if the person need rehabilitation. This phase is commonly misunderstood because many people see rehabilitation as part of the secondary prevention in which only medical care professionals provide services to rehabilitate the affected individual, but in general, going beyond the medical model, rehabilitation is more than just medical care, many times it means, helping the affected individual to improve the quality of life and to come back to society in a productive manner.

In addition, the following table goes into more details and provide examples that should help to understand the above mentioned concepts:



## Levels of Prevention

| Major concepts | Major concepts |
|---|---|
| **Primary care,** is the first level of entry into the health care system: the visit to the family physician, ER, clinic | **Primary prevention,** involves halting any occurrence of disease before it happens: health promotion, health education, and health protection |
| **Secondary care,** is usually given in the hospital, extended care in a nursing home, or home health care: minor/major surgery, routine/advanced nursing care | **Secondary prevention,** includes health screening and detection activities used to discover pathogenic states in individuals within a population with the intention to block the progression of disease |
| **Tertiary care,** is the care found in larger, advanced-care hospitals which use the highest levels of technology: open heart surgery, brain surgery, and specialty intensive care units | **Tertiary prevention,** intents to retard or block the progression of a disability, condition, or disorder: rehabilitation services, patient education, after care, health counseling |

Image prepared by Giovanni Antunez, Licensed CC BY 4.0

**The levels of prevention can be linked to the clinical model,** and this helps to put into context the common medical and public health interventions that most people are used to be seen as part of the work of the medical and public health work. See a summary of this comparison in the following table:

| In the clinical model, there are three levels of health care services | From the clinical model, three level of prevention have emerged |
| --- | --- |
| Primary care | Primary prevention |
| Secondary care | Secondary prevention |
| Tertiary care | Tertiary prevention |

### Important comments about the information above

The above distinctions should help public health professionals in training, especially health educators to better understand the differences between the levels of care and the level of prevention, especially for students who are not familiar with the United States healthcare system. This system works differently in other parts of the world, especially in developing countries. So, the comparison should also help to conciliate the differences.

### Summary of the chapter.

This chapter has intended to give the student among other things, an overview of the field of epidemiology, what is it, and how it works. Also, it had covered most common concepts used in epidemiology; most of these concepts will be integrated with the content of several of the chapters that made the book.

# 2. History of Epidemiology

*Learning Objectives*

By the end of this chapter, the learner will be able to

- Learn about the history of epidemiology, which parallels the history of public health
- Examine the history of epidemiology as a relatively new science
- Illustrate the history of epidemiology with events that not only happened in the traditionally European tradition, but also in the United States, and the Americas.
- Create a history of the field of epidemiology and public health in the United States.

**Epidemiology: classic historic events**

The following content intents to provide only the most basic historic events in epidemiology. It does not intent to be exhaustive, but at least it will covered the most salient events that had influenced the development of what it is known now as the field/science of epidemiology. From ancient civilizations, the time of Greeks and Romans, and the 17th, 1800s, 1990s to the 20th and 21st century the main purpose of epidemiology and public health has been the search for determinants that can explain causation and disease, the discovery of treatments and moreover, the prevention and control of disease, and related health problems.

**Medicine/Public health in the time of the Greeks**

Although epidemiology is considered a new science, in reality even before it was recognized as a science, it probably started by the times of the Greeks, who attributed disease causality to environmental factors influenced by the gods, the presence of epidemics and deadly toll of diseases were influenced by those factors. Also, at the time of the ancient Greeks, disease was also explained by what is known as, the theory of the four ‘fluids’ (blood, phlegm, yellow bile and black bile) that is in its turn premised on the theory of the four elements (fire, earth, water and air) and their four corresponding qualities (heat, dry, humid and cold).[1]


Image of 'the concepts of the four fluids/four elements'. Public Domain. Wikipedia.

**Early sanitation efforts, water systems and toilets**

Then, the Romans, who developed aqueducts and sewage systems in order to prevent to prevent the development of diseases related to the lack of proper sanitation; bring water to their cities improved sanitation, and personal hygiene, which are essential for good health.[2]

1. Yannis T. (2009). The historical origins of the basic concepts of health promotion and education: the role of ancient Greek philosophy and medicine, Health Promotion International, 24 (2), 185–192. From https://doi.org/10.1093/heapro/dap006
2. Deming D. (2020). The Aqueducts and Water Supply of Ancient Rome. Ground water, 58(1), 152–161. https://doi.org/10.1111/gwat.12958

'Pont du Gard', By Benh LIEU SONG , CC BY-SA 3.0, Wikimedia.

It is the Romans who also developed the first rudimentary system of disposition of **excretas** before what we know as toilets were invented.[3]

'Ostia Toilets'(Ancient Roman latrines / latrinae), Public Domain,Wikimedia commons.

### Medieval period: The Black Death (or, Plague)

3. Gill, N.S. (August 02, 2019). Roman Baths and Hygiene in Ancient Rome. ThoughtCo. From https://www.thoughtco.com/hygiene-in-ancient-rome-and-baths-119136

Then, the medieval period with the presence of deadly diseases such as the plague.[4]

'Citizens of Tournai bury plague victims,' by Pierart dou Tielt, Wikimedia commons, Public Domain.

### The Renaissance

With the advent of the sciences and arts during the Renaissance period, the initial work on medicine, and anatomic studies started. Several individuals flourished during those times, bringing new ideas, and discoveries including Girolamo Fracastoro (1478–1553), an Italian physician who introduced the term '**fomites**' still in use today in Epidemiology to refer to those inanimate objects that might contribute indirectly to the transmission of an infectious disease. And of course, the anatomical draws and notes done by Leonardo DaVinci were part of the contributions from this era.[5]

4. 'Black Death,' Wikipedia, licensed Creative Commons Attribution-ShareAlike License 3.0 URL: https://en.wikipedia.org/wiki/Black_Death
5. MNT Editorial Team. (Feb. 8, 2022). What was medieval and Renaissance medicine? Medical News Today (MNT). From https://www.medicalnewstoday.com/articles/323533

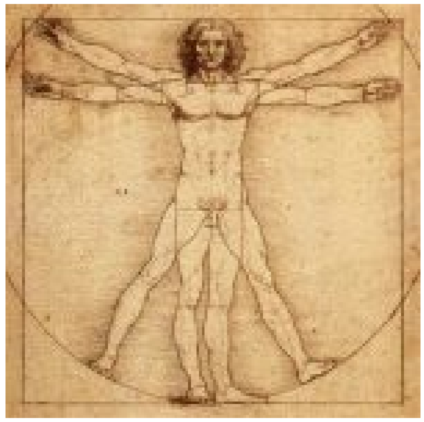


Image of 'The Renaissance man' by Leornardo DaVinci. Public Domain.

**Medicine without microscope**

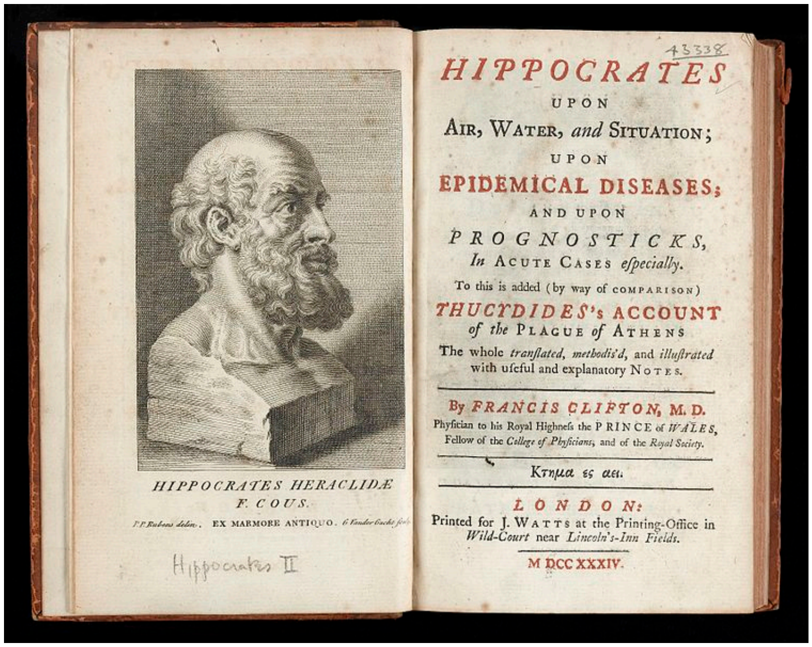
If history of medicine and epidemiology goes back to the beginning, there is a common factor for the majority of events (that make history), they revolved around the search for **causation**, and how to prevent and treat disease. Since the majority of deaths in earlier times of the human race were caused mostly by infectious agents (usually bacteria), the progress in identifying these causes had to wait until the discovery of the microscope since most of these agents cannot be seen with the naked eye.[6]

6. Poppick, L. (March 30, 2017). Let Us Now Praise the Invention of the Microscope In Smithsonian Magazine. Available at: https://www.smithsonianmag.com/science-nature/what-we-owe-to-the-invention-microscope-180962725/

‘Old-microscopes’. Image by Idal. Wikimedia Commons.
Licensed CC BY-SA 3.0 Unported

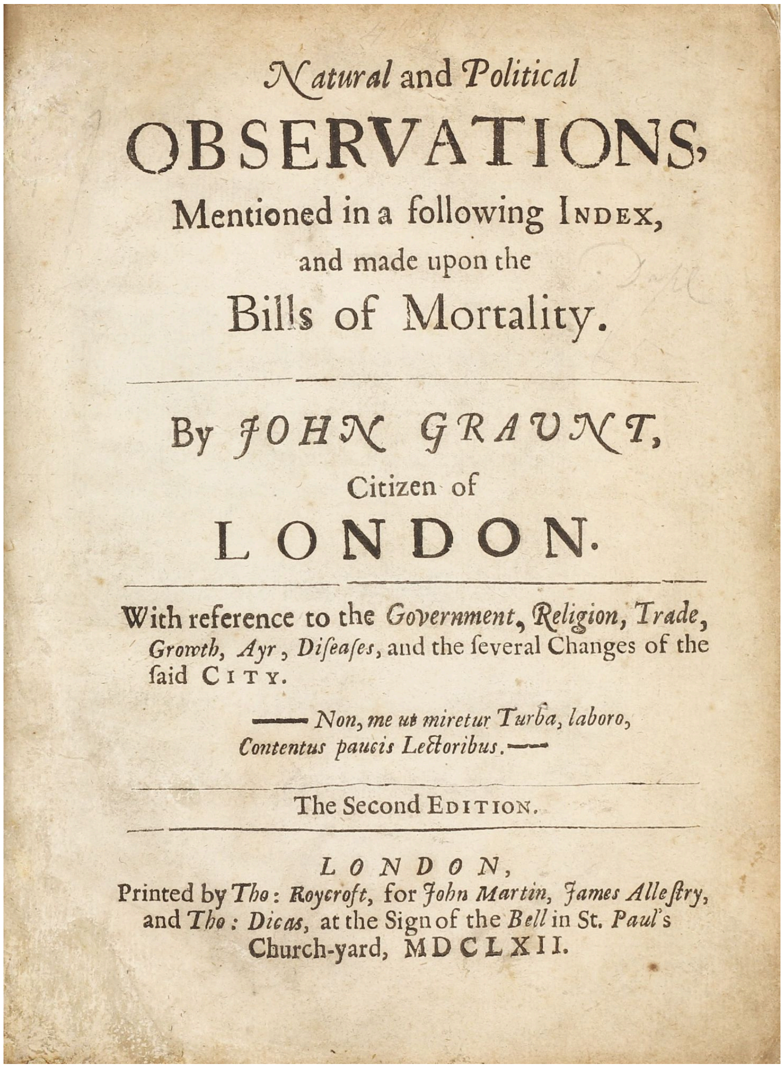
In this context, the first systematic methods used to identify causation are attributed to the father of medicine, **Hippocrates** who in about 400 years of the contemporary era was able to propose that there is a connection among three major elements of nature and health or disease, these are: *air, water, and place* (as we will see later in the book, *place* is usually referred as location – where people reside/live); he proposed for example that in some cases the ‘bad air’ can cause disease, or, the ‘quality of the waters’ as he called, and the place in which people live such as high altitudes and other geographical characteristics, especially the soil, and terrain.[7]

7. Centers for Disease Control and Prevention (CDC). (n.d.). Lesson 1: Introduction to Epidemiology Section 2: Historical Evolution of Epidemiology. Circa 400 B.C. section. Available at https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section2.html




'Hippocrates bust and title page', image from
Wikimedia Commons. Licensed CC BY 4.0 International.

Another important person in the history of epidemiology is, John Graunt, a London cloth merchant who lived in the middle of the 17th century. Due to his work, he is considered the first demographer and initiator of the concept of vital statistics.**John Graunt** In 1603 in London, Graunt compiled the first register of births and deaths in England, under the name of, the 'Bills of Mortality'. This is consider, the first systematic recording of deaths known in the history of record keeping and reporting of deaths, and other vital events. Graunt's intentions were to show that human life conforms certain predictable statistical patterns, he wrote for example that deaths varied recorded age, sex, who died, of what, and where they died (location), and when (time of death).[8,9]



Natural and Political
OBSERVATIONS,
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.

By JOHN GRAUNT,
Citizen of
LONDON.

With reference to the Government, Religion, Trade, Growth, Ayr, Diseases, and the several Changes of the said CITY.

——Non, me ut miretur Turba, laboro,
Contentus paucis Lectoribus.——

The Second EDITION.

LONDON,
Printed by Tho: Roycroft, for John Martin, James Allestry, and Tho: Dicas, at the Sign of the Bell in St. Paul's Church-yard, MDCLXII.

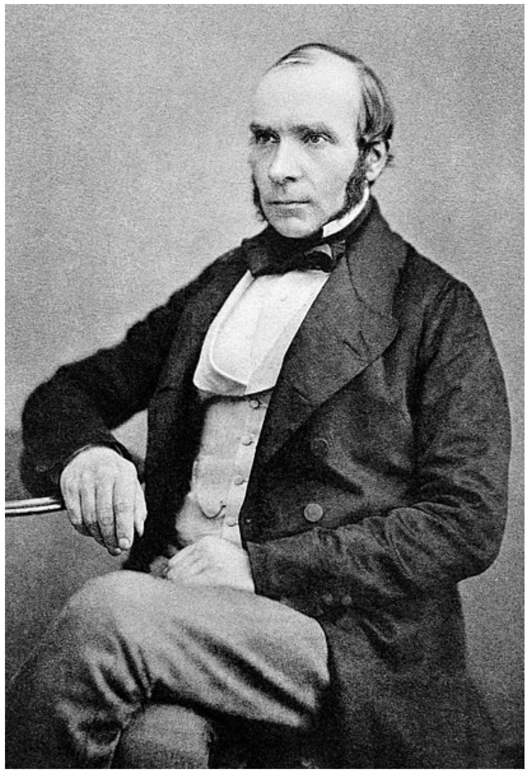
'The Bills of Mortality', Image from Linda Hall Library. Public Domain.

Joh

In the other in which events are presented here, the next person who made significant contributions to epidemiology and public health is **Dr. John Snow** –

8. Stephan, E. (n.d.). John Graunt (1620-1674), Observations in the Bills of Mortality. From http://www.edstephan.org/Graunt/graunt.html
9. Linda Hall Library. (2020, April 24). Scientist of the Day - John Graunt. From https://www.lindahall.org/john-graunt/

already discussed in another section of the chapter. One of the main remarks is that **Dr. John Snow** is considered the father of public health.



Dr. John Snow. Image from Wikimedia Commons. Public Domain.

Dr. John Snow (1813-1858) was a respected physician, and the anesthesiologist to Queen Victoria of England, he studied cholera throughout is medical career, from his studies he established sound and useful epidemiological methodologies. He used several research methods that were highly advanced for this time, but also methods that are still valid today. Using a 'natural experiment' approach Dr. Snow studied an epidemic of cholera that develop in 1849 in London (already discussed previously in this chapter).


Dr. Snow's map, image from Wikimedia Commons. Public Domain.

He developed a famous map, that at that time was considered 'scandalous' because it was not drawn to scale but it provided an operational system in which he was able to identify and registered where persons with cholera lived and worked, the distribution of cases were mapped using a **dot map** (this is considered an incipient beginning of the geographic information systems **GIS** map known as

**choropleth map** ), the collected information was used by him eventually identified that the cause of the cholera epidemic was a problem with contaminated drinking water from one of the main source of water in London. Through his interventions, Dr. Snow was able to stop the cholera epidemic.[10], [11].

**Cholera in the United States**

During the winter of 1848-49, an outbreak of cholera was reported in New Orleans, Louisiana. A disease that had killed thousands of people in Europe had also affected the U.S. population in the South. This came after a brief epidemic of cholera in 1832. According to an article published in the famous local newspaper, The Times Picayune, the source of the outbreak came from the contagion of locals with the arrival of a passenger ship that was bringing immigrants from Liverpool and Le Havre in December 22 of 1948. In a period of two months, the outbreak took the lives of approximately 3,000 New Orleans residents who were not prepared to deal with the contagion. The local newspaper did not realized that the cause of the disease was already know as an illness caused by contaminated water with the Vibrio Cholera bacterium; however some locals had noted that most people affected by the outbreak were low income individuals who live in improper sanitary conditions, and who may have no access to clean drinking water. The cholera outbreak mentioned above has several similarities with the epidemic described by Dr. John Snow in England in the 1849, people affected were those with low access to health education, and proper sanitation. Actually, the affected local resident were reported as people of “irregular habits,” in the same article some numbers are also reported that some of victims of cholera were also people from Ireland, Germany, England, Spain, Prussia, France and Switzerland.[12]

10. Kneisl, K. (February 9, 2021). John Snow, A Turning Point In Epidemiology. From https://storymaps.arcgis.com/stories/1913fb6e17cd49c88b801e4c6edb67bf
11. University of California Los Angeles (UCLA). (2003). Who is John Snow? UCLA Dept. of Epidemiology, Fielding School of Public Health. From http://www.ph.ucla.edu/epi/snow.html
12. No author. (April 23, 2020). "Careless of Cholera": The New Orleans Outbreak of 1848. The Times Picayune. From https://www.pitothouse.org/blog/2020/4/23/careless-of-cholera-the-new-orleans-outbreak-of- 1848

‘The Duke of Orléans visiting the sick at the Hôtel-Dieu hospital, during the cholera outbreak, in 1832.’ Painting by Alfred Johannot, 1832. Licensed as Public Domain, from Musée Carnavalet.

**A response to the Cholera epidemic: The creation of a school of public health in New Orleans**

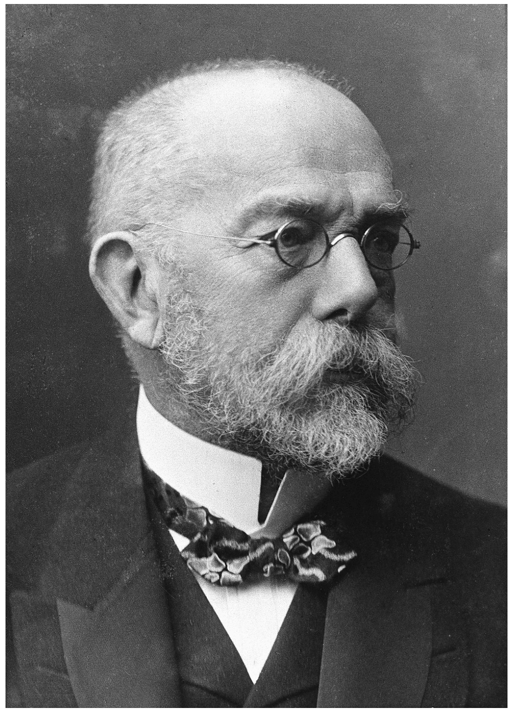
As a student in the school of public health at Tulane university, New Orleans, Louisiana, I heard so many times from professors that problems such as the cholera epidemic in 1832 were one of the reasons to create one of the first public health schools in the U.S., the Tulane School of Public Health and Tropical Medicine founded in 1834; other similar public health problems were also part of the reasons , also as part of its origins. These other health issues malaria, yellow fever, and smallpox; all of them affected the region and a group of physicians founded a medical college in New Orleans (not yet known as the School of Public Health) so training on these issues and because of the great need to education of tropical medicine problems in Louisiana and other parts of the South.[13]. See below an old (historical) picture of New Orleans:

13. No author. (2020). History of Tulane SPHTM, a Timeline. From https://sph.tulane.edu/timeline

“The City of New Orleans and the Mississippi River c1885” Licensed Public Domain, Library of Congress.

**The discoveries of Dr. Robert Koch and its importance in public health**

It is probably insufficient to have a section on historical events, or, individuals and their works without mentioning **Dr. Robert Koch** (1843-1910), German physician and one of the main founders of bacteriology as a science, among other great accomplishments he discovered the anthrax disease cycle and the bacteria responsible for tuberculosis and cholera. He received the Nobel Prize for Physiology or Medicine in 1905 for his research on tuberculosis.



Dr. Robert Koch, image from Wikimedia Commons.
Public Domain.

Dr. Koch's work has been pivotal for the work on causation since most of the infectious diseases in his time were bacteria, or, bacilli. He developed what is called, the Koch postulates that include the following four criteria are: (1) The microorganism must be found in diseased but not healthy individuals; (2) The microorganism must be cultured from the diseased individual; (3) Inoculation of a healthy individual with the cultured microorganism must recapitulated the disease; and finally (4) The microorganism must be re-isolated from the inoculated, diseased individual and matched to the original microorganism. These postulates are the founding for the reproducibility of experimentation with infectious agents, and the search for health determinants.[14].

**Connecting the information above**

How is that Dr. Koch was connected to say to the history of the cholera epidemic mentioned above? The answer is that if Dr. Snow would have access to a microscope, and also had the bacteriological knowledge that was available in the times of Dr. Koch, including his own experiments with bacteria, then, Dr. Snow

14. Segre J. A. (2013). What does it take to satisfy Koch's postulates two centuries later? Microbial genomics and Propionibacteria acnes. The Journal of investigative dermatology, 133(9), 2141–2142. From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3775492/

would not say that the contaminated water (in the time of his assessment of the cholera epidemic in London, 1849) contained a great amount of “impurity in the form of small white, flocculent particles...[visible to the naked eye]. [15] but to the presence of the Vibrio cholerae bacterium. So, although separated by decades of difference, the connection exists  between Dr. Snow studies and Dr. Koch discoveries of the presence of bacilli in the contaminated water.[16]

**The foundations of the concept of vital statistics in public health**

Another individual who made a difference in the field of epidemiology is, **William Farr,** a British epidemiologist who was appointed compiler of abstracts in England, 1839. His work examined the linkage between mortality rates and population density, it also created the foundation for the concept of vital statistics. Farr’s contributions provided foundation for classification of diseases (**ICD system**).[17]

15. Snow, J. (1855). On the Mode of Communication of Cholera. From http://www.ph.ucla.edu/epi/snow/snowbook_a2.html
16. Rogers HR.(1895). Dr. Robert Koch and His Germ Theory of Cholera. JAMA, XXIV(23):903–904. doi:10.1001/jama.1895.02430230037013
17. Lilienfeld, DE. (2007). Celebration: William Farr (1807–1883)—an appreciation on the 200th anniversary of his birth. International Journal of Epidemiology, 36(5), 985–987. From https://academic.oup.com/ije/article/36/5/985/775018

'William Far', image from Wikimedia Commons. Public Domain.

**Systematization of the science of epidemiology**

And, coming back to home in the United States,  and according to most historians in the field of epidemiology and public health, *the first epidemiologist* in the United States is **Dr. Wade Hampton Frost**, who in 1921 was appointed professor of epidemioogy at Johns Hopkins University School of Hygiene & Public Health, Baltimore, Maryland.

Dr. Wade Hampton Frost, image from Wikimedia Commons. Public Domain.

Based on what it has been presented above, this means the beginning of the teaching of epidemiology in a public health school in the U.S., he advocated for the use of quantitative methods to support the public health research that at that time was mainly research on the study of infectious diseases.[18]

Another individual who made great contributions to epidemiology and public health in the United States, **Dr. Alexander Langmuir,** another great scientist who established CDC's Epidemiologic Intelligence Service (EIS).

18. Johns Hopkins Bloomberg School of Public Health. (December 20, 2012). Medical Historian Speaks at School about Wade Hampton Frost (web article). From https://publichealth.jhu.edu/2004/frost

Dr. Alexander Langmuir, image from Wikimedia Commons. Public Domain.

Dr. Langmuir did extraordinary contribution to not only the CDC as an institution, but also to public health in general. From 1949, he served as first Chief Epidemiologist at CDC and remained in that position until 1970. Prior to this, in 1952, he convened the first Conference of State and Territorial Epidemiologists, which became the organization representing epidemiologists in the United States. Overall, Dr. Langmuir defined disease surveillance at CDC.[19]

**Note:** All of the above individuals made great contributions to the the field of epidemiology and public health. The names of more individuals could continue, but the main focus of the chapter at this point will continue with the establishing of parallel between the main influence of in public health and epidemiology which is the British system. It is important to recall that the first school of public health in the world is the London School of Hygiene and Public Health, which more a long time, was the only school in the world, hence its main influence in the United States public health system.

### The history of epidemiology in the U.S. and England

What I am proposing here is that in the U.S., we start writing about our own history

19. Thacker, SB. (n.d.). Epidemiology and Public Health at CDC. Office of Workplace and Career Development, Office of the Director. From https://www.cdc.gov/mmwr/preview/mmwrhtml/su5502a2.htm

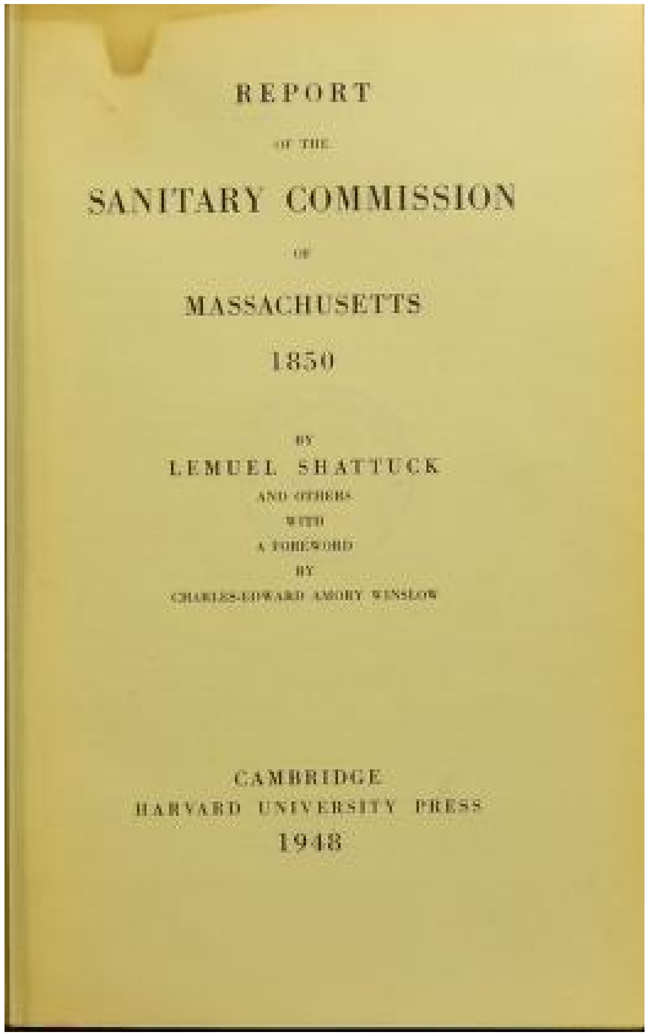
of epidemiology, not because the historical events of Europe (especially in England) don't matter (because they do, they are important milestones in the history of public health in general, and  epidemiology); but because there is need to have historical information that applies to our times, and needs of the population we served in public health.

**Parallel:**

**"The Shattuck Report" in the U.S. and the progress of public health in England**

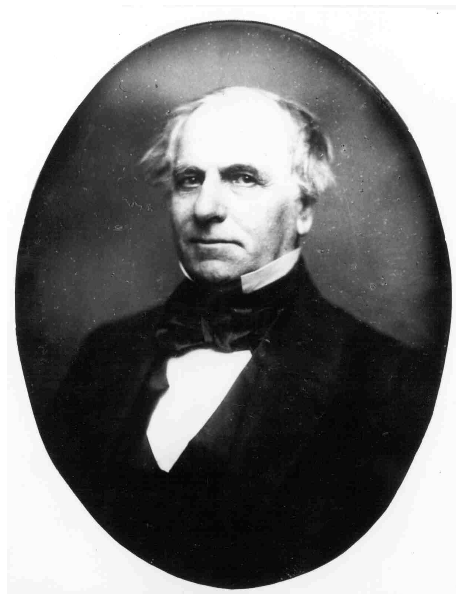
The history of public health in the United States also saw an event that marked the beginnings of  what is essentially the beginning of public health in the U.S., this event is the publication of the Shattuck Report, written by Dr. Lemuel Shattuck and published in 1850 by the Massachusetts Sanitary Commission, it reports on the serious health problems experienced by certain groups of people, who were also living in unsanitary conditions in the city of Boston. Among the recommendations given to solve these problems, the report recommended the creation of state health department and local boards of health in each town so public health interventions to address the issues were developed and implemented. [20],[21]

20. Report of the Sanitary Commission of Massachusetts 1850. Report available in pdf from https://www.google.com/search?client=firefox-b-1-d&q=the+shattuck+report
21. Winkelstein, W. (July 2008). Lemuel Shattuck Architect of American Public Health. Epidemiology, 19(4), 634. From https://journals.lww.com/epidem/Fulltext/2008/07000/Lemuel_Shattuck__Architect_of_American_Public.21.aspx




Dr. Shattuck, image from London School of Hyg, & Trop. Med.

As a result (of the Shattuck's report) in 1866 in New York City for the first time in the United States an organization as the report recommended was created. This is considered by many as the beginning of public health in the U.S. A picture of Dr. Shattuck is shown below:



Dr. Shattuck, image from [London School of Hyg, & Trop. Med.]

These early beginnings of public health in the U.S. contrasted so much with the reality of public health in England and France that were more concern on how to properly administer and manage the incipient health care system already in place. In other words, the U.S. at that time (1850) were dealing with issues that were a thing of the past in England, or, at least something that they have under control already.[22] Again, this brief summary of the impact of the Shattuck's report is another

22. Britannica. (n.d.). National developments in the 18th and 19th centuries. From https://www.britannica.com/topic/public-health/National-developments-in-the-18th-and-19th-centuries

example that this country could learn from the public health lessons that England had already experienced.

**Parallel Dr. John Snow and Cholera (in England) and Seat belts use in the U.S.**

For example, Dr. John Snow (1884) in England, used epidemiological methods to identify the cause (origin) of the cholera epidemic that was killing in thousands; and also developed the strategy to stop the epidemic, which was basically removing the pump's handle of the contaminated water source; in this manner, people would stop drinking the contaminated water, and the epidemic stopped.[23] See image of the pump below:



**Seat belts law in the U.S.**

A similar example can be found in the U.S. when the use of seat belts started as a public health measure to save lives during motor vehicle accidents, in the early 1970s when the measures were implemented, most people did not buckle up; they thought it was optional; but with the pass of time people started seeing the benefits of seat belt use.[24] When the regulations are enforced, people who drive without a seat belt would get a traffic ticket from the police; and just because getting tickets is not pleasant, but also, just by knowing that the ticket is recompense for not wearing a seat belt, has made the measure very effective. And, actually, it was probably the

23. University of California Los Angeles (UCLA). (2003). John Snow and the Broad Street Pump. UCLA Dept. of Epidemiology. From https://www.ph.ucla.edu/epi/snow/snowcricketarticle.html
24. Centers for Disease Control and Prevention. (n.d.). Policy Impact: Seat Belts. From https://www.cdc.gov/transportationsafety/seatbeltbrief/index.html

best way to make people use (consistently) seat belts while driving; and nowadays, it has become a habit, a norm.[25].

*Seat Belts Have Saved an Estimated 255,000 Lives Since 1975*

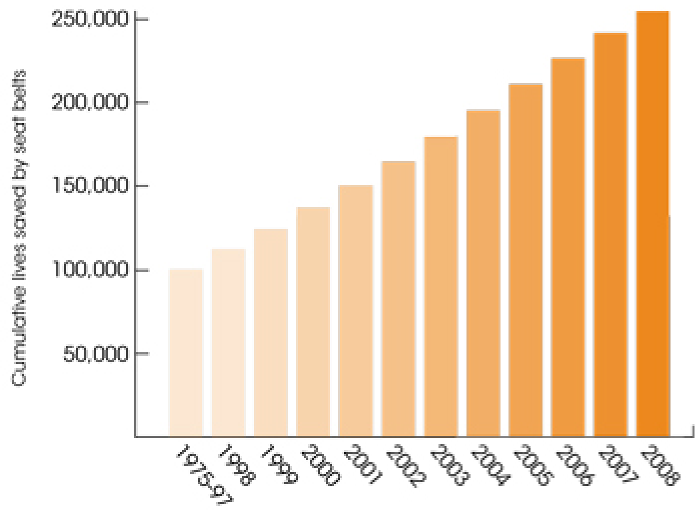



Image from [CDC](). [Public Domain]().

**Lesson learned:** During Dr. Snow's times in England (1884) the epidemic stopped with he removed the water pump handle that delivered the contaminated water, in this manner people stopped drinking from a contaminated water source. Similarly in the United States, when the use of car's seat belts stopped the growing number of motor vehicle accidents when the seat belt law took place in the year 1968. People needed to change behavior, but without removing barriers is hard for people to change. Laws help to enforce public health policies.

**Another parallel between England and the United States, the second wave of immigrants in the U.S.**

Most books in epidemiology used the historical events of public health in England in order to draw a parallel with the U.S. history of public health; and this works well until we realize that the history of public health, and epidemiology in the U.S.

25. Dept of Transportation (US), National Highway Traffic Safety Administration (NHTSA). (2010). Traffic Safety Facts: Seat Belt Use in 2010—Overall Results. Washington (DC): NHTSA. From http://www-nrd.nhtsa.dot.gov/Pubs/811378.pdfpdf icon

has to do with one major issue, the wave of immigrants from distant countries, and how these immigrants dealt with the current sanitary conditions, housing, and environment conditions of their time, especially those immigrants that decided to reside in New York; which was not prepared to receive such high numbers of people, especially the unsanitary conditions of the buildings that many of these immigrants were allocated in the city. crowdedness, lack of proper disposal of excretas, and high level of cross-contamination of food, and water were typically in a for example Irish immigrant household (say, apartment) in New York. These immigrants faced a very difficult time keeping themselves healthy and alive (especially their children, who in many cases died of dysentery, or, any other type of oral-fecal related disease.[26]

"New York, NY, yard of tenement (between 1900 and 1910)" Image from Library of Congress. Public Domain.

These above mentioned conditions made the epidemiological landscape a very much different from those people living in Europe (just say, in England) were more epidemiological research was conducted at that time. So, the U.S. epidemiology in this case was a little different – probably many methodological similarities, or, very

26. Library of Congress. (n.d.). Adaptation and Assimilation. Presentation, Immigration and Relocation in U.S. History. From https://www.loc.gov/classroom-materials/immigration/irish/adaptation-and-assimilation/

different in the sense that the U.S. was experience a public health crisis while Europe was already ahead of the U.S. in the area of sanitary controls, and housing conditions, to mention some of the public health issues at that time.[27]

*Remarks*

In this book, the history of public health in England is used to correlate with the U.S. public health (and epidemiological) history, which had several similarities, but at the same time, different because what affected the population in the U.S. was a thing of the past in Europe. Most events used to write the history of epidemiology in Europe occurred later in this new world that we know now as the Americas, which includes the United States.[28]

**Next parallel: tobacco use and cardiovascular disease as a public health problem in England and the United States**

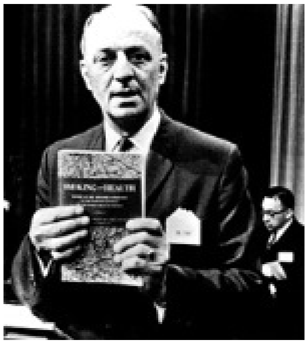
Another historical event (or, series of events) in the history of epidemiology and public health is the **tobacco use and its effects on health, England versus the United States**

Another parallel between the history of epidemiology, and public health and England, and the U.S. is the case of the tobacco use and heart disease. The early regulations started earlier in England, and the English people started hearing about the noxious effects of tobacco use (mainly smoking cigarettes) way before people in the U.S., and it was not until 1957 when this country (the U.S.) became aware of the need to study (do some research) the problem, and considered for the first time to start regulations on its use, one of the main event is the released of the U.S. Surgeon General's Advisory Committee on Smoking and Health report (on January 11, 1964) that concluded that cigarette smoking is a cause of lung cancer and laryngeal cancer in men, a probable cause of lung cancer in women, and the most important cause of chronic bronchitis.[29].

Although now, it seems something not of such a major importance, it represented

27. Markel, H., & Stern, A. M. (2002). The foreignness of germs: the persistent association of immigrants and disease in American society. *The Milbank quarterly*, *80*(4), 757–v. https://doi.org/10.1111/1468-0009.00030 From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7152289/
28. Comments of the author.
29. Centers for Disease Control and Prevention. (n.d.). History of the Surgeon General's Reports on Smoking and Health. From https://www.cdc.gov/tobacco/data_statistics/sgr/history/index.htm

a big step in the search for health determinants at that time. Medicine also benefited from these studies, because now, physicians, and other health care providers have data from their own people about the effects of chronic tobacco use and cardiovascular disease.



Picture of the 1964 U.S. Surgeon's General Luther L. Terry, M.D., who released the first report on the nocebo effects of smoking on health. By CDC. Public domain.

**Another historical public health/epidemiology events in the United States: Research on Cardiovascular Disease**

In the search for the risk factors associated with Cardiovascular disease three major cohort studies have been conducted, these are known as, The Framingham Heart Study in Massachusetts, The Bogalusa Heart Study in Louisiana, and the San Antonio Heart Study in Texas. Each one of these studies will be briefly described in the following paragraphs, and more detailed information will be also provided in the chapter on study designs, section on cohort studies.[30]

**The Framingham Heart Study**

It represents one of the greater milestones in epidemiology of cardiovascular diseases studies in the U.S., and it also became well-known later in the world. The Framingham Heart Study (FHS) in Massachusetts, it took its name from a town in eastern Massachusetts, U.S. which became selected as the site of the study.[31] See a picture of the Framingham City Hall below:

30. See cohort studies section in chapter 5.
31. Bacallao Gallestey, J. (n.d). Framingham Heart Study research project, Framingham, Massachusetts, United States. Encyclopedia Britannica. From https://www.britannica.com/event/Framingham-Heart-Studay

"Town hall in Framingham, Mass." Encyclopedia Britannica. EagleOne GNU Free Documentation License Version 1.2

The Framingham Heart Study is the first cohort to study the U.S. population in the issue of heart disease, and its related risk factors. It started in 1948 with a small group of individuals, and it is still collecting information about the white population in the U.S., and heart disease behavioral lifestyle risk factors. The findings of the Framingham Heart Study provided much of the common knowledge that people now have about heart disease, including the effects of exercise, diet, and smoking on cardiovascular disease.[32]

### The Bogalusa Heart Study

Later in the history of epidemiology in the U.S., there is another major studies focused on cardiovascular disease; but this time on the black population, its name the Bogalusa Heart Study, Bogalusa is a city (mainly black population) in Washington Parish, Louisiana, United States. It is the principal city of the Bogalusa Micropolitan Statistical Area, which includes all of Washington Parish and is also part of the larger New Orleans–Metairie–Hammond combined statistical area.[33]
See Louisiana Map below:

32. No Author. (n.d.). Framingham Heart Study, Participant cohorts. Available at https://www.framinghamheartstudy.org/participants/participant-cohorts/
33. Bogalusa Louisiana. From Wikipedia, https://en.wikipedia.org/wiki/Bogalusa,_Louisiana

Bogalusa Louisiana. From Wikimedia Commons. Licensed CC 3.0

Bogalusa became well-known because of this study, which began in 1972 being the main investigator, Dr. Gerald S. Berenson, a Bogalusa native and pediatric cardiologist who recognized the need for investigations into the childhood antecedents of adult cardiovascular diseases.[34] I had the privilege to met Dr. Berenson while doing my public health doctoral studies at Tulane University in New Orleans. [35]. To see additional images of the city of Bogalusa, see below:

"Bogalusa City Hall". Wikimedia Commons. Licensed CC BY-SA 4.0

The main contributions of the Bogalusa Heart Study among others include the following: adult heart disease, atherosclerosis, coronary heart disease, and essential hypertension begin in childhood, cardiovascular risk factors can be identified in early life, but the levels of risk factors in childhood are different than those in the adult

34. No Author. (n.d.). Bogalusa Heart Study. Available at: https://www.clersite.org/bogalusaheartstudy/
35. Personal comments from the author.

years. In addition, another great contribution of the Bogalusa Heart Study among others is that it is targeted (collected data) to African American children (while the Framingham targeted mainly adults).[36]

The next step, **the San Antonio Heart Study (SAHS)**
With the past of time, the need to have information (data) about cardiovascular disease among another major ethnic group in the U.S. became one of the major reasons to conduct another famous cohort study on the Latino/Hispanic population in the U.S., the name of the study, *San Antonio Heart Study, San Antonio, Texas.* See a picture of San Antonio below:

"Downtown San Antonio." Wikimedia Commons. CC BY-SA 3.0

The San Antonio Heart study is an ongoing study that focuses not only on cardiovascular disease, but also in the study of type II diabetes on Latinos residing in the U.S. The author could not find a specific web site for the San Antonio Heart Study but the following is a journal article that provides substantial information about it.[37]

My personal critique/comments of the San Antonio Heart study is that it lacks representativity of the different U.S. Latino ethnic groups, by being mainly focused on Mexican-American, which constitute a great majority of the population of Latinos in the U.S. and excluding the great numbers of new immigrants from other countries, and regions especially Central American and the Caribbean, the data results cannot be generalized to the rest of the U.S. population in the U.S., but on a positive side, the contribution of the San Antonio Heart Study (SAHS) cannot

36. Tulane Center for Lifespam Epidemiology Research (CLER). (n.d.). Bogalusa Heart Study. Available at: https://www.clersite.org/bogalusaheartstudy/
37. Shen D, Mitchell B, Hazuda H, Clark G, and Stern M. (1992). The San Antonio heart study research information study, Proceedings Computers in Cardiology, 607-610. Available at: https://ieeexplore.ieee.org/document/269385/keywords#keywords

be denied because it represents pioneering research on two major problems (cardiovascular disease and type II diabetes) of morbidity, and mortality among U.S. Latino population.

*Key Takeaways*

As we had presented above, the study of cardiovascular disease, and its associated factors represent one of the major milestones in epidemiological investigation in the U.S. and a history of the field of epidemiology without looking at these studies will be short; and that is probably one of main reasons for including them in this historical section of the book.

In addition to those parallels covered above, one more issue I want to review is the issue of research in the field of epidemiology in the U.S. and public health. As it has been seen in the examples above, **the major research efforts on the topic of cardiovascular disease in the U.S. has been a reflection of the systemic problem on race**, why some ethnic groups have been excluded from some studies? Why some ethnic groups have been lied, and abused in some research studies? In the name of sciences and medical advances, abuses have been made. The following is an example of this type of problem.

**Medical and Public Health Research going wrong in the U.S., the case of the Tuskegee syphilis study**

The Tuskegee syphilis study (actually called at that time, the 'Tuskegee Study of Untreated Syphilis in the Negro Male.') took place in Macon county, Alabama during the years of 1932 and 1972, during this time a study of the natural history of syphilis was conducted among the African American population in the area. Without person's informed consent, the clients in the study were left without treatment for syphilis, and were offered regular medical care for other health problems but not for syphilis. They were also offered free meals, and burial insurance.[38]

38. CDC. (n.d.). The U.S. Public Health Service Syphilis Study at Tuskegee. From https://www.cdc.gov/tuskegee/index.html

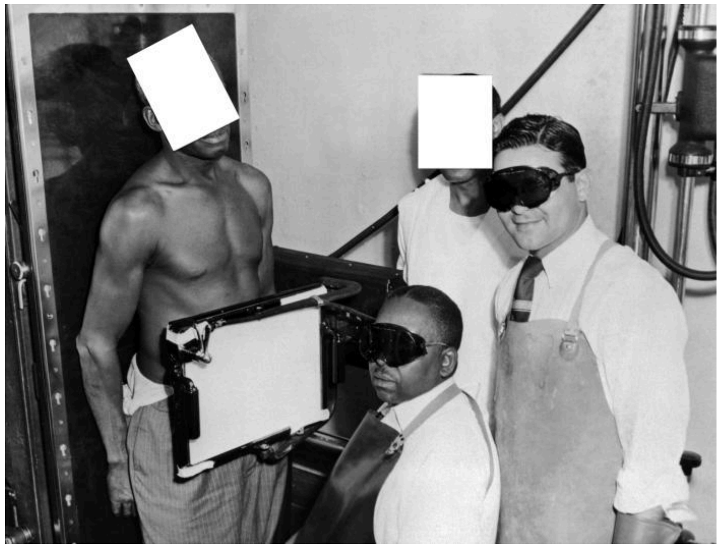


Image from [CDC Public Health Image Library (PHIL).](#)

According to the CDC this photograph was taken around 1932, it shows participants in the Tuskegee Syphilis Study. In this picture, an African-American man was being x-rayed, while in the standing position.

At the beginning of the study, the number of participants included 600 African American men, from this group, 399 had syphilis and 201 did not have the disease. During the time of the study, the investigators told these men that were being treated by "bad blood," a common term at that time that was used to describe some common diseases such as syphilis, anemia and fatigue. The study lasted for 40 years, although the men participating in the study we told that the study would last 6 months.[39]

The Tuskegee syphilis study is an example of human experimentation of of the clinical history of a common infection disease, which by the time of the study there was a treatment for syphilis already available. The withholding of the treatment exposed the study participants, their families, and generations of African Americans

39. National Archives of Atlanta. (n.d.). Tuskegee Syphilis Study. From https://www.archives.gov/atlanta/exhibits/item470-exh.html

in the south. For this reason, this study is frequently cited as an example of what an investigator is not supposed to do in conducting research, especially research on human subjects. The question that usually comes to mind when the author reads about this study is, what can be done to prevent similar situations in the future of research in the U.S.? And, the question remains open.

**The epidemic of smallpox in Europe and then, later in the United States**

From the list of plagues that affected the world population in the past, one of the major problem is smallpox, a disease now **eradicated**. A disease caused by the variola virus . It caused millions of deaths especially during the Medieval times, although it is believed to have existed for at least 3000 years. [40]The person who created the first smallpox vaccine is Dr. Edward Jenner, an English doctor who as part of his observations noted that milkmaids who had gotten cowpox were protected from smallpox, which means that they received immunity (this concept was unknown at his time) from having the disease, this became the base for the development of the smallpox vaccine. Having a vaccine made a tremendous difference in the history of the disease, which basically contributed to its eradication. [41], [42]

40. World Health Organization (WHO). (No date). Smallpox. From https://www.who.int/health-topics/smallpox#tab=tab_1
41. Riedel S. (2005). Edward Jenner and the history of smallpox and vaccination. Proceedings (Baylor University. Medical Center), 18(1), 21–25. https://doi.org/10.1080/08998280.2005.11928028
42. CDC. (No date). History of Smallpox. From https://www.cdc.gov/smallpox/history/history.html


"Edward Jenner (1749–1823), Public Domain, from the National Library of Medicine

**Smallpox in the United States – a terrible experience for Native Americans**

Smallpox was mainly a European problem for a long time, until the colonization of the new world (America) in which the colonizers brought the disease to the natives, which did not have immunity since they were never exposed to the smallpox agent. So, smallpox killed thousands of natives (not only in what is known as North America today, but also in the Caribbean and South America. One of the worst example is the killing of Native Americans in the United States when the colonizer gave (as a gift, at least once) blankets intentionally contaminated with smallpox – this is also an example of an early episode of terrorism using biological weapons, as it was expected the strategy worked, and again entire populations of Native Americans were killed by the disease. [43]

43. Kiger, PJ. (Nov 25, 2019). Did Colonists Give Infected Blankets to Native Americans as Biological Warfare? There's evidence that British colonists in 18th-century America gave Native Americans smallpox-infected blankets at least once—but did it work? History, 15. From https://www.history.com/news/colonists-native-americans-smallpox-blankets

“Treaty of Penn with Indians by Benjamin West”, Public Domain

Wikimedia Commons.

**Influenza Pandemics in history**

The history of epidemiology and public health has been shaped by the presence of pandemics: plague, flu and recently, coronavirus or, COVID-19. Through history, these pandemic have been responsible for the death of thousand of people around the world. Overall, what had changed from ancient times to today’s public health is the availability of vaccines, and also the improvement in water and sanitation around the world, with major gains for developed nations who have better systems of drinking water, and sanitation. Until recently, the fear for a major pandemic has been for a worldwide scale flu that could kill thousands of people. But instead of the flu, the world received the coronavirus pandemic, which is still around. Since flu has been the major fear before coronavirus, this section will cover some of the major events of flu pandemics.

**Flu pandemics**

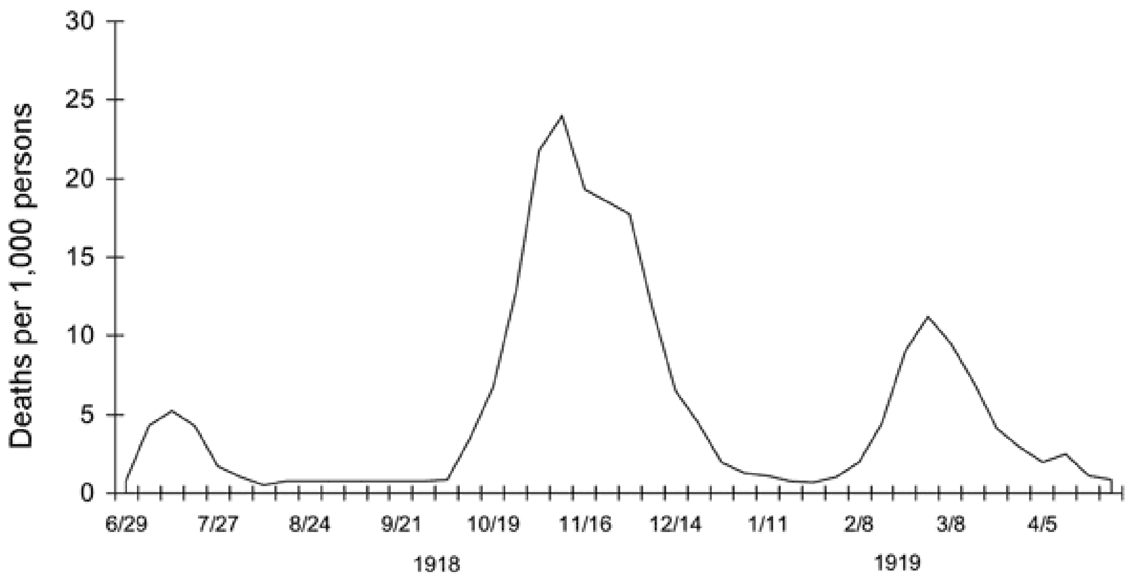
In general, flu pandemics,  had repeated in history every 100 years, because of that the ‘expected’ and ‘predicted’ pandemic was on flu (and not coronavirus).[44].

44. Kertscher, T. (April 10, 2020). Fact-check: Has a pandemic occurred every 100 years? PolitiFact. From https://www.statesman.com/story/news/politics/elections/2020/04/10/fact-check-has-pandemic-occurred-

Because of its importance in history, the first major pandemic that appears usually in most history books, is the flu pandemic.

**The flu pandemic of 1918**

The 1918 Influenza (flu) Pandemic (also called, the Spanish flu, because of the great mortality in that country at that time). This famous pandemic lasted from 1918-1919 and killed 50 to 100 million persons worldwide. The pandemic had three waves as it is represented in the following image:




"1918 spanish flu waves," Image Licensed Public Domain, Wikipedia.

As it is seen in the graph/image above, the second wave of the 1918 flu pandemic was higher/stronger than the first, and the third wave was also higher than the initial wave. For this reason, the 1918 flu pandemic was called, "The mother of All Pandemics,"[45] of course this designation has been before the current coronavirus pandemic. There was also a 1927 flu pandemic that it was less noticed around the world, but it killed significant number of people, especially army soldiers in the United States.[46]

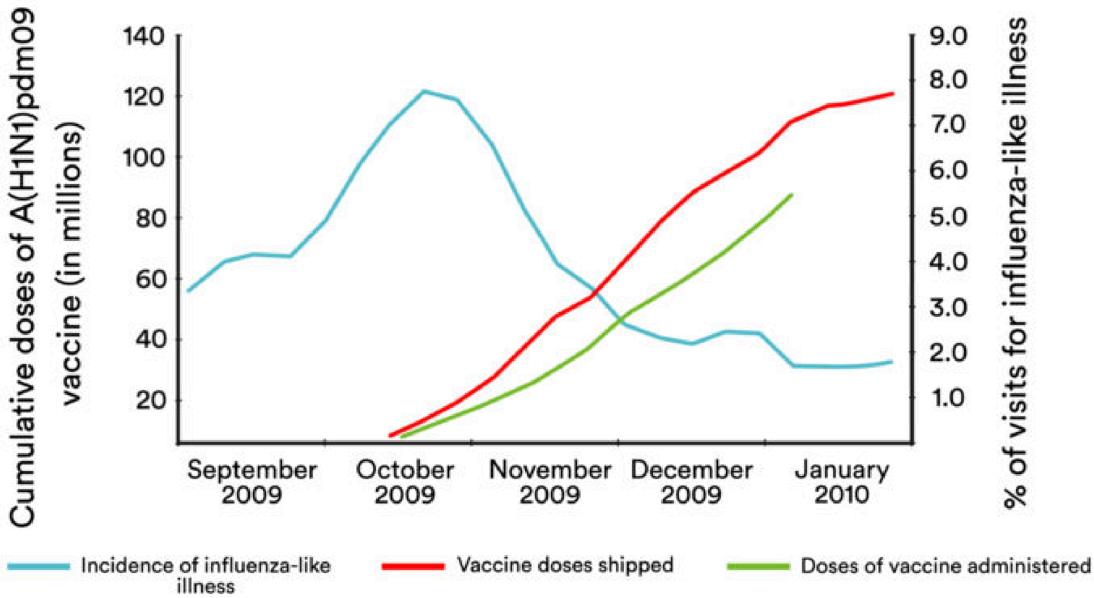
**The 2009 H1N1 influenza pandemic: a repeat of the 1919 pandemic**

every-100-years/984128007/

45. Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: the mother of all pandemics. Emerging infectious diseases, 12(1), 15–22. https://doi.org/10.3201/eid1201.050979
46. Barry, J. M. (2020). The great influenza: The story of the deadliest pandemic in history. Penguin UK.

The specific time of this pandemic is from 2009-2010, the disease was first identified in the United States, and eventually was named as the 2009 H1N1 influenza. The first two cases of the diseases were reported by the Centers for Disease Control and Prevention (CDC) in April 2009, after this, the number of cases grew to 60 million by summer 2010. The pandemic also attacked several other countries in the world, and it was known to have a preference for the group of people 14-64 years, with those been 65 and older as the less affected group.[47], [48] This pandemic should had served as a lesson for the U.S., but a lot of skepticism grew that the possible for another major pandemic was not possible, or, if this would happen then, the country had enough public health resources including vaccines to control the problem, which has been proven wrong with the current coronavirus pandemic making manifest that the U.S. public health system was not prepared for a pandemic.

47. CDC. (no date). 2009 H1N1 Pandemic (H1N1pdm09 virus). From https://www.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html
48. Mayo Clinic Staff. (Feb 24, 2021). H1N1 flu (swine flu). From https://www.mayoclinic.org/diseases-conditions/swine-flu/symptoms-causes/syc-20378103



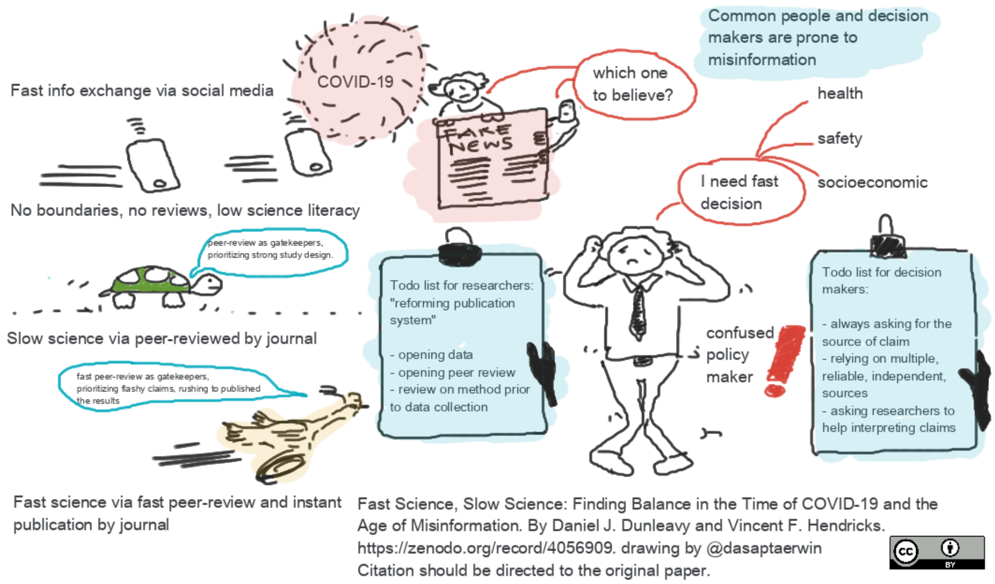

"Comprehensive vaccine initiative report," Image by, from CIDRAP Comprehensive Influenza Vaccine Initiative report [49], [50]

**The most recent pandemic: the COVID-19 pandemic due to Coronavirus 2019-current**

The prediction of a new pandemic was always there but the medical and public health preparations were never in place, so, the pandemic started in Wuhan, China with an amount of infected persons never seen before, the agent a coronavirus named COVID-19 because of the year in which it was reported as the cause of the pandemic. The news were everywhere, but the United States did not pay attention to the situation, until almost two months later in March 2020; acknowledging and responding to the pandemic was proven bad, because the country and the world did not have a vaccine for it, nor the U.S. had the public health infrastructure to respond properly, there was a lot of unknown about the disease, and especially its

49. Branswell, H. (June 11, 2019). The last pandemic was a 'quiet killer.' Ten years after swine flu, no one can predict the next one. STAT, Health. From https://www.statnews.com/2019/06/11/h1n1-swine-flu-10-years-later/
50. Osterholm, MT., Kelley, NS., Manske, JM., Ballering, KS., Leighton, TR. (October 2012). CIDRAP Comprehensive Influenza Vaccine Initiative report. University of Minnesota, Academic Health Center. From https://www.cidrap.umn.edu/sites/default/files/public/downloads/ccivi_report.pdf

form of transmission; the measures taken since the beginning did not have a social acceptance in a society that values so much its personal freedom, and having the federal government, and public health authorities, and experts tell the public to wear a mask, stay at home and related mandates created an epidemic of misinformation, and contradictory messages that have been hard to eliminate until today.[51],[52]. I have found an image -shown below that summarizes some of the reasons for this campaign of misinformation that has been one of the major problems of the COVID-19 pandemic.





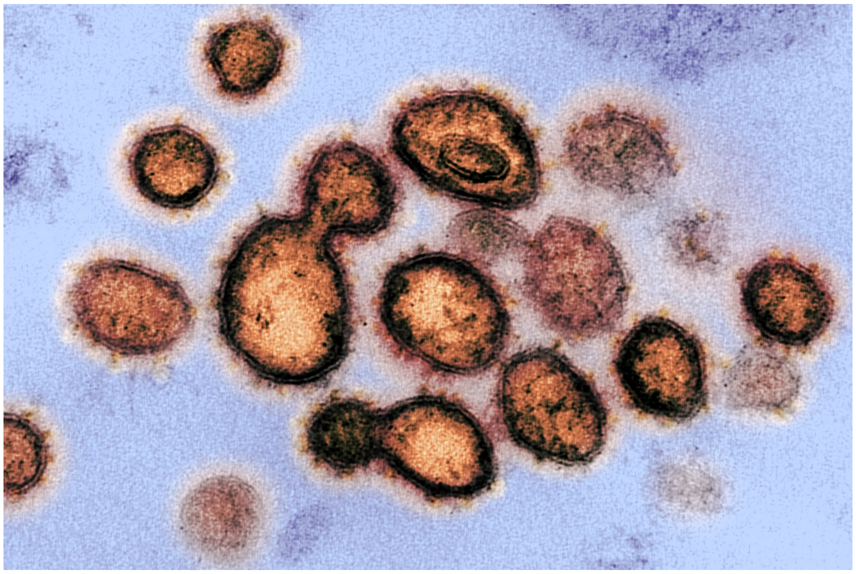
There is also an article that refers to more details about the content of the image.[53]

### The agent, and the statistics

As mentioned before the agent of the coronavirus that has many different variants,

51. Tagliabue,F.,Galassi,L., Mariani,P.(2020). The "Pandemic" of Disinformation in COVID-19. SN Comprehensive Clinical Medicine, 2 (9),1287. From https://doi.org/10.1007/s42399-020-00439-1
52. Grimes, DR. (March 12, 2021). Medical disinformation and the unviable nature of COVID-19 conspiracy theories. PLOS One (Pone). From https://doi.org/10.1371/journal.pone.0245900
53. Dunleavy, Daniel, & Hendricks, Vincent. (2020, September 28). Fast Science, Slow Science: Finding Balance in the Time of COVID-19 and the Age of Misinformation (Version 1). Zenodo. http://doi.org/10.5281/zenodo.4056909

including those variants that cause the respiratory illness, called COVID-19 or, SARS-CoV-2 that may have been originated in an animal (still not clear which type), and the virus mutated causing the illness in humans.[54]



"Novel Coronavirus SARS-CoV-2 (or, COVID-19) electronic microscope image," Licensed Public Domain,, from rawpixel.

The numbers of infected people had been high since the beginning of the pandemic with an excess mortality, which has been the main feature because before an infected individual died, the need for hospitalizations had been never seen in decades, and probably in the last century. The COVID-19 pandemic had officially affected 191 countries and territories in the world. Most of the data collected ranges from 2020-2021.[55] Recent estimates and reports still only available in websites such as one of the most popular, which is from the Johns Hopkins Coronavirus Resource Center.[56]. Since the estimates are still changing, I decided to present a map of the world distribution of the pandemic to illustrate the content.

54. Smriti Mallapaty, S. (28 January 2022). Where did Omicron come from? Three key theories The highly transmissible variant emerged with a host of unusual mutations. Now scientists are trying to work out how it evolved. Nature. From https://www.nature.com/articles/d41586-022-00215-2
55. Haidong Wang, H., Paulson, KR., Pease, SA., Watson, S. et al. (April 16, 2022). Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. The Lancet, 399 (10334), 1513. From https://doi.org/10.1016/S0140-6736(21)02796-3
56. No author. (No date). Johns Hopkins University Coronavirus Resource Center. From https://coronavirus.jhu.edu/map.html

"Global distribution of estimated excess mortality rate due to the COVID-19 pandemic, for the cumulative period 2020–21"
Image from an article by Haidong Wang, H., Paulson, KR., Pease, SA., Watson, S. et al.[57]

At the time I completed this chapter, more and more information about the COVID-19 pandemic became available in the world wide web, scientific medical journal, and gray literature especially from popular magazines, or, open educational resources. For now, the information above had been written with the intention to have an idea of the history of public health and epidemiology through out two-three centuries.

**Summary**

Although a historical review of events, and individuals, it shows that the history of public health and epidemiology are intertwined. The historical events mentioned in this chapter confirm that epidemiology as a science has made great contributions to medicine and public health, in the search for determinants and also in the design of effective public health interventions.

57. aidong Wang, H., Paulson, KR., Pease, SA., Watson, S. et al. (April 16, 2022). Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. The Lancet, 399 (10334), 1513. From https://doi.org/10.1016/S0140-6736(21)02796-3

# Timeline of history of epidemiology

This is a  timeline of the major historical events mentioned in the "History of Epidemiology" chapter:

*An interactive H5P element has been excluded from this version of the text. You can view it online here:*
*https://minnstate.pressbooks.pub/hgantunez/?p=1137#h5p-1*

# 3. Epidemiology of Infectious Diseases

*Learning Objectives*

By the end of this chapter the reader will be able to:

- Learn about the field of infectious diseases as the initial field of study in epidemiology
- Learn about the epidemiological triangle and its related terms commonly use in the area of infectious diseases
- Relate prevention of disease outbreaks to modes of transmission
- Analyze the concepts of infectivity, pathogenicity, incubation period
- Apply the learned concepts to real life situations, especially during epidemics

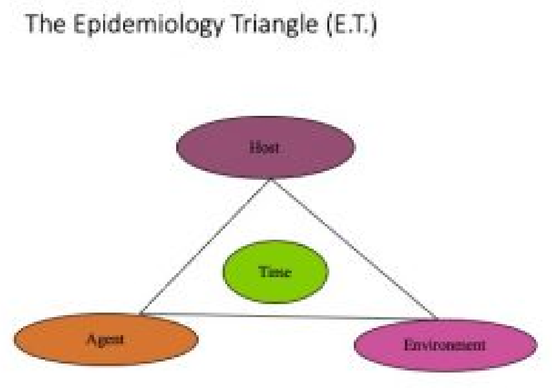
**Introduction**

Epidemiology basically started as a science for the study of infectious diseases. Why infectious diseases? It is because since the beginning of times, infectious diseases have been the main cause of mortality in the world. For this reason, it is said that infectious diseases created epidemiology as a science that study health phenomena, especially disease, and mainly transmissible diseases. Also, it can be said that without infectious diseases, there is no epidemiology.[1] That is the reason for this chapter to exist ahead of the majority of topics covered in the rest of this textbook.

**How things got started?**

Everything started with the epidemiologic triangle and its components – already covered at the beginning of this book. I am posting the image again to refresh the concept. See below:

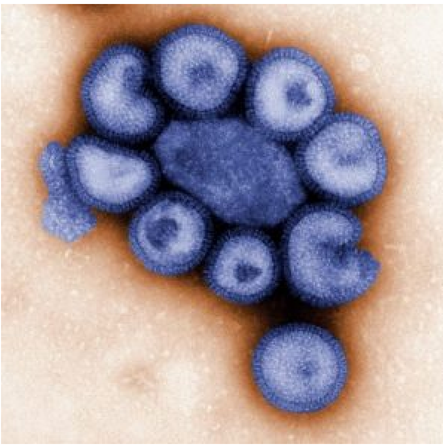
1. Barreto, M. L., Teixeira, M. G., & Carmo, E. H. (2006). Infectious diseases epidemiology. Journal of epidemiology and community health, 60(3), 192–195. https://doi.org/10.1136/jech.2003.011593




In this model, for disease to occur, there is need to have a *host*, an *agent*, and an *environment*.[2] The variable, *time* was added, later. In this manner, the interaction among host, agent and environment happen in a certain period of time.

**The agent** (also, called, infectious disease agent, or, **pathogen**
It is accepted that the agent is one of the major factors, without an agent there is no disease [because this is biologic/medical model]. Common agents include: bacteria, viruses, fungi, other microbes and parasites.[3], [4] Examples of common agents include, the viruses. See picture below:

2. CDC. (n.d.). Lesson 1 Understanding the Epidemiologic Triangle through Infectious Disease, pdf material. From https://www.cdc.gov/healthyschools/bam/teachers/documents/epi_1_triangle.pdf
3. Mayo Clinic. (n.d.). Types of infectious agents. From https://www.mayoclinic.org/diseases-conditions/infectious-diseases/multimedia/types-of-infectious-agents/img-20008643
4. Janeway, CA., Travers, P., Walport, M. et al. (2001). Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science. Infectious agents and how they cause disease. From: https://www.ncbi.nlm.nih.gov/books/NBK27114/



Electronic picture of a virus. James Gathany, Judy Schmidt, USCDCP on Pixnio" data-url="http://James Gathany, Judy Schmidt, USCDCP on Pixnio">Photo by Judy Schmidt, Image from PIXNIO. Public Domain.

### The host

The host can be a person, or an animal that harbors the disease.[5] As described later in the textbook, a host can show symptoms of the disease, or, can be free of symptoms.

#### The environment

In general it is accepted that the environment is everything that surrounds the host, or, promote the existence of the agent.[6] It is is the external (to the host or, agent) set of factors that contribute to the development of the disease. **Time** It refers to the duration of the disease, it includes incubation periods (not symptoms are manifested during this period), or, the manifestation of the whole set of symptoms that characterize the disease.

### Disease Transmission

In addition to the definitions above, it is customary to review additional concepts that help to explain how infectious diseases occur. The first of this is, transmission.

#### Common ways of transmission

**Transmission** goes together with the concept of chain of infection [also visited before in this book]. See image below about the general process of transmission.

5. Bowden, S. E. & Drake, J. M. (2013) Ecology of multi-host pathogens of animals. Nature Education Knowledge 4(8):5. From https://www.nature.com/scitable/knowledge/library/ecology-of-multi-host-pathogens-of-animals-105288915/
6. Gupta, A., Gupta, R., Singh, R.L. (2017). Microbes and Environment. In: Singh, R. (eds) Principles and Applications of Environmental Biotechnology for a Sustainable Future. Applied Environmental Science and Engineering for a Sustainable Future. Springer, Singapore. https://doi.org/10.1007/978-981-10-1866-4_3

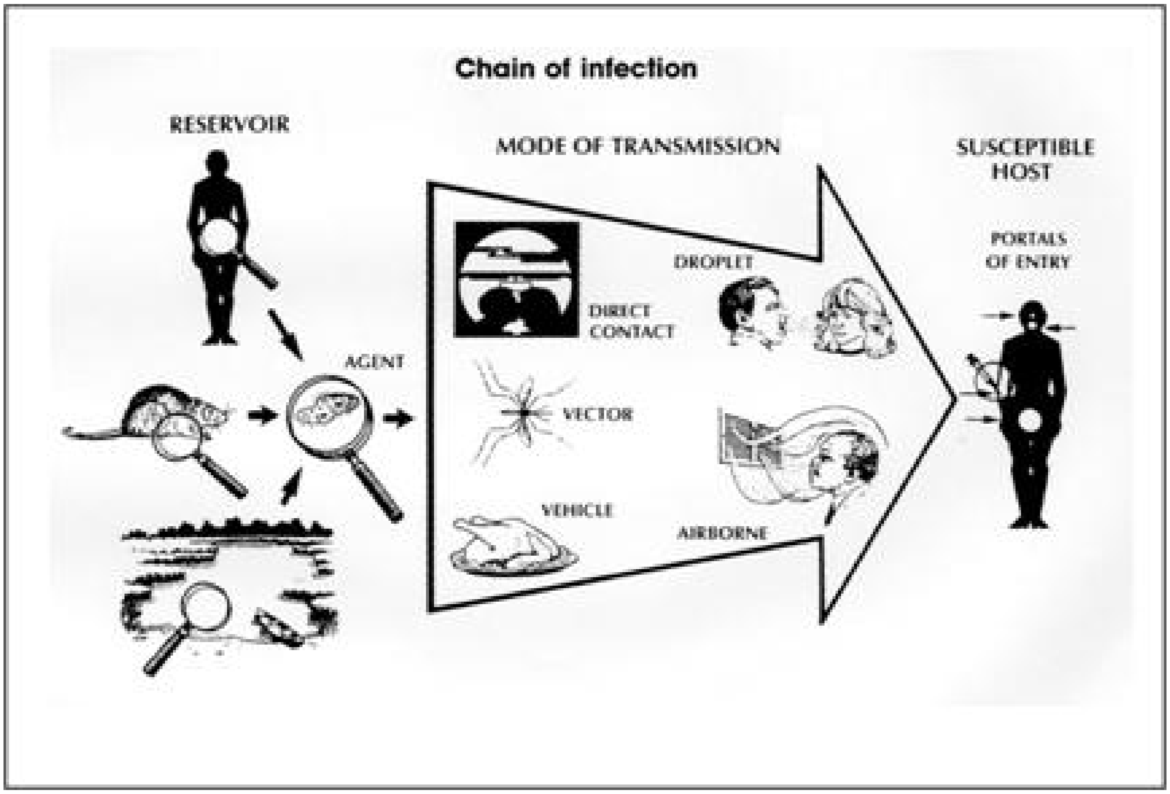



Image of The chain of infection. CDC. Public Domain.

In this context, diseases can be transmitted directly or, indirectly. **Direct transmission** is the spread of disease from person to person. Examples, kissing, touching and **infected person** (or, the person who has the disease). It can occur through direct physical contact, or direct person-to-person contact, such as touching with contaminated hands, skin-to-skin contact, kissing, or sexual intercourse. And, in other situations, the direct transmission occurs while a person is coughing or, sneezing in which the **droplets** (of infection) spread from an infected person to a susceptible host**susceptible host**[7]

**Indirect Transmission**, in this case, the spread of the infection occurs through an **intermediate** source called, vehicles, fomites, or, vectors. These terms require the use of some definitions starting with **vehicle**, which is the medium that contains the infection agent. Examples of vehicles include used needles contaminated with

7. Valencia, H.V. (July 26, 2017). How Are Diseases Transmitted? Healthline. From https://www.healthline.com/health/disease-transmission

blood, which is common among IV substance users, or, in the case of needles accidentally contaminated with hepatitis. Another common example of a vehicle is contaminated water.

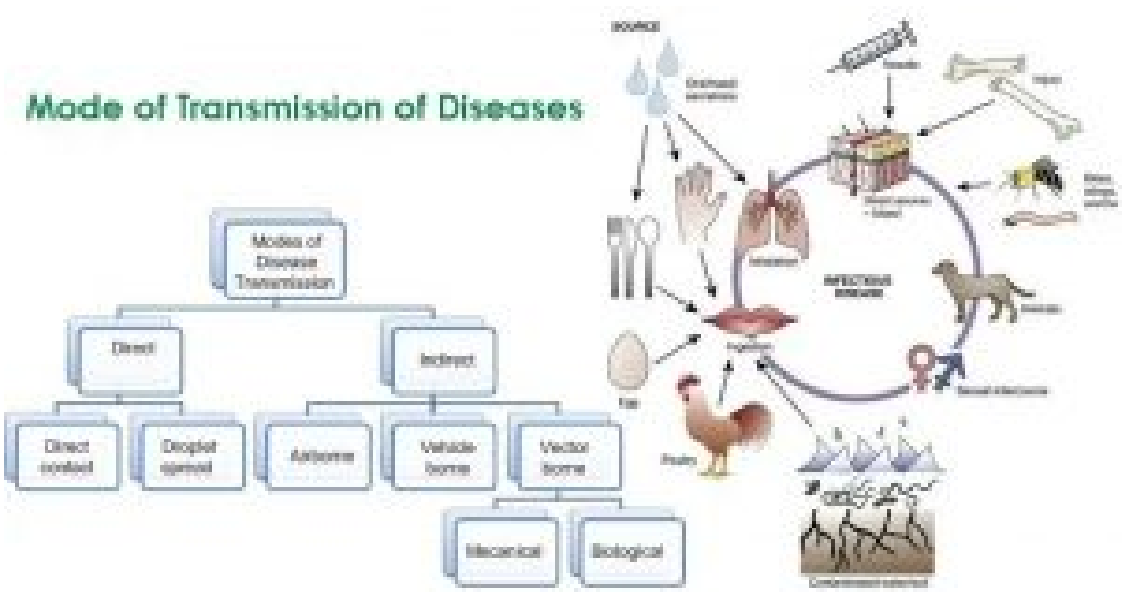
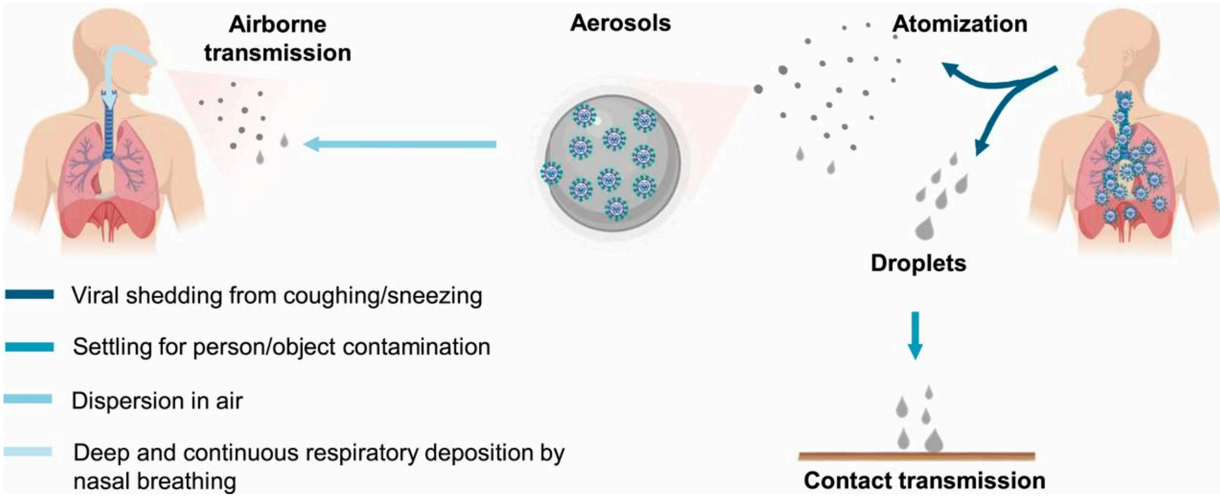



Image Source: [Indian Journal of Health Sciences and Biomedical Research KLEU](#) and [Pinterest](#)

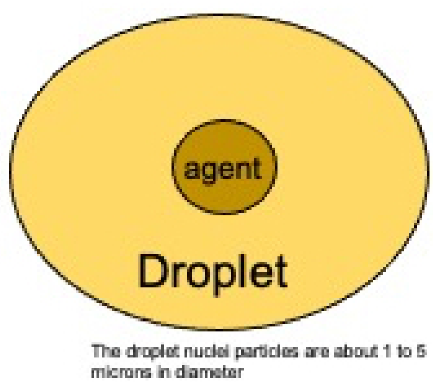
Other forms of Indirect transmission, **airborne**, in this situation, the disease is spread via droplets when a person sneezes, coughs, or talks, spraying microscopic pathogens. It can also happens when dust particles are spread by fans in abandoned buildings.

The mechanism for **airborne indirect transmission.** Example, when an infected person coughs or sneezes, tiny particles containing droplet nuclei are expelled into the air.




'Airborne transmission,' image from Wikimedia, Licensed CC BY 4.0 International.

Since its importance, the concept of **droplet nuclei** is defined here, which is a particle 1-5 micron in diameter that contains the infectious agent inside (see image below). The droplets nuclei can remain suspended in the air for several hours, depending on the environment. The most effective droplet nuclei tend to have a diameter of ≤ 5 μm.[8]




'The Droplet nuclei', Image prepared by Giovanni Antunez, Licensed CC BY 4.0

To have an idea of the generation of droplets' nuclei, it helps to compared how many

8. Atkinson J, Chartier Y, Pessoa-Silva CL, et al. editors. (2009). Natural Ventilation for Infection Control in Health-Care Settings. Geneva: World Health Organization. Annex C, Respiratory droplets. From: https://www.ncbi.nlm.nih.gov/books/NBK143281/

of these droplets can be generating while coughing, talking, or, sneezing. See image and comments below:

'Sneezing,' image from CDC PHIL, Public Domain.

- In general it is accepted that one cough can generate 3000 droplet nuclei.
- Talking for 5 minutes can generate 3000 droplet nuclei and singing can generate 3000 droplet nuclei in one minute
- Sneezing generates the most droplet nuclei by far (tens of thousands), which can spread to individuals up to 10 feet away.[9]

Ways in which **Indirect transmission** occurs. Examples include, waterborne, vehicleborne, foodborne, and vectorborne. Details of these forms of transmission are discussed in more detail below:

**Waterborne,** this type of indirect transmission is when a pathogen (for example, shigella, cholera) is carried via drinking water, swimming pools, streams, lakes. This type of transmission is more frequent for example in some periods of the year, for example summer when most people are more active and visit recreation centers that have public pools, or, when people visit lakes.

9. Tang, J. W., Li, Y., Eames, I., Chan, P. K., & Ridgway, G. L. (2006). Factors involved in the aerosol transmission of infection and control of ventilation in healthcare premises. The Journal of hospital infection, 64(2), 100–114. https://doi.org/10.1016/j.jhin.2006.05.022

Shigella' bacteria, image from Medicinenet.

**Vehicleborne**, in this case it is related to fomites (eating utensils, clothing, washing items, combs, etc.). Typical examples include, scabies and, lice (head and pubic).

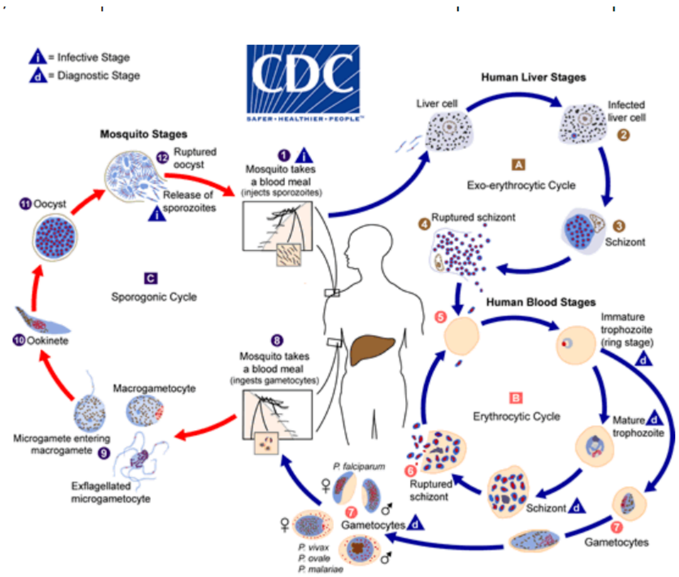
'Sarcoptes scabei' image from Wikimedia.

'Pediculosis capitis', the cause of a head lice, image from Flickr.

'Pediculosis pubis' the cause of Pubic lice, from Wikimedia.

**Vectorborne**, although redundant, the disease is transmitted by a vector such as mosquitoes, ticks, etc. For some these are considered simple mechanical processes, as when the pathogen uses a host (a fly, flea, lice, or rat) as a mechanism for a ride, for nourishment, or as a physical transfer process in order to spread -this process is called, **mechanical transmission**.

In the same manner, when the pathogen undergoes changes as part of its life cycle, while within the host/vector and before being transmitted to the new host, it is called, **biologic transmission**. This process is easily seen in malaria, in which the female Anopheles mosquitoes blood meal is required for the Plasmodium protozoan to complete its sexual development cycle. This can occur only with the ingested blood nutrients found in the intestine of the Anopheles mosquito. See the process in an image below:




[‘Malaria Cycle,’](#) Image from [CDC,](#) [Public Domain](#).

### Additional concepts in the transmission of an infectious disease

#### Reservoirs

This term refer to humans, animals, plants, soil, or inanimate organic matter (feces or food) in which infectious organisms live and multiply. For example, the reservoir for *Clostridium tetani* that causes tetanus is

‘Soil’, image by Jayme Burrows at The Spruce.

commonly contaminated soil. This information is very important especially for individuals who do gardening, or, work in occupations that require handling of soil and have open wounds that could be contaminated with the Clostridium spores.[10]

**Fomite(s)** which is inanimate objects that may become contaminated with the infectious agent. A common example are contaminated diapers in a day care center. Diapers are also fomites that may contribute to the spread of some infectious agents, especially, gastrointestinal and respiratory infections.[11]

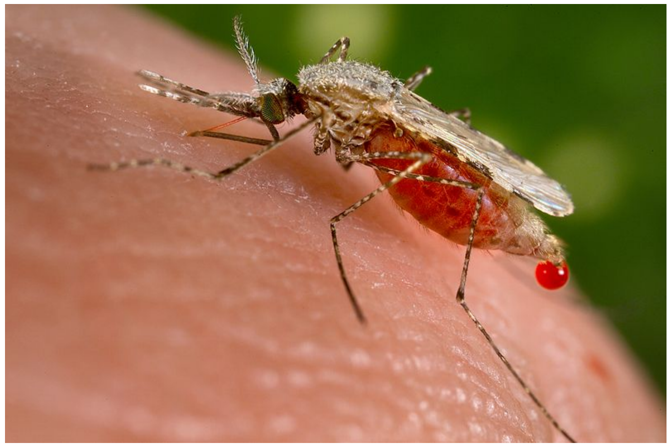
Image from Diapers.

Another example is personal protective equipment (PPE) used for decontamination. Although fomites seem important in the transmission and spread on an infectious disease, fomites are not effective (they not always contribute) to the transmission of disease.

The last of the terms in the topic of indirect transmission is **vectors**, it refers usually to insects and small animals that contribute to the spread of disease as part of their life cycle. Common examples of vectors are fleas, mites, flies, mosquitoes, small rodents, etc. These vectors contribute to the spread of disease in an

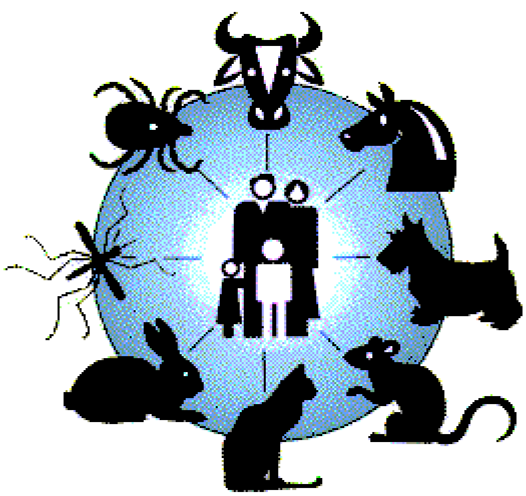
10. George EK, De Jesus O, Vivekanandan R. (2022, Jan). Clostridium Tetani. [Updated 2021 Aug 30]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing. From https://www.ncbi.nlm.nih.gov/books/NBK482484/
11. Boone, S. A., & Gerba, C. P. (2007). Significance of fomites in the spread of respiratory and enteric viral disease. Applied and environmental microbiology, 73(6), 1687–1696. https://doi.org/10.1128/AEM.02051-06

effective manner, so, their control is part of the preventive measures of well-known infectious diseases such as malaria, zika virus, west nile virus and others.



Common Disease Vector, the mosquito that transmit malaria, image from Wikimedia, Public Domain.

**Additional disease transmission concepts** that relate to the concept model of the Epidemiological Triangle (E.T.). The following are these concepts:



'Zoonosis,' Image from Texas Health Dept.

**Zoonosis**, when an animal transmits the disease to a human. Examples of this type of diseases include, malaria,lice, tinea, dengue fever, and others.

There is also, the term **Zoonoses**, those diseases and infectious transmitted between vertebrate animals and man (or, humans). Examples include, mad cow disease, equine meningitis, cryptosporidium, hantavirus, toxoplasmosis, rabies, and others.

**Enzootic**, this term is not commonly used, and it refers to a disease that only affect animals, small number of them, and in a persistent manner (or, endemic).

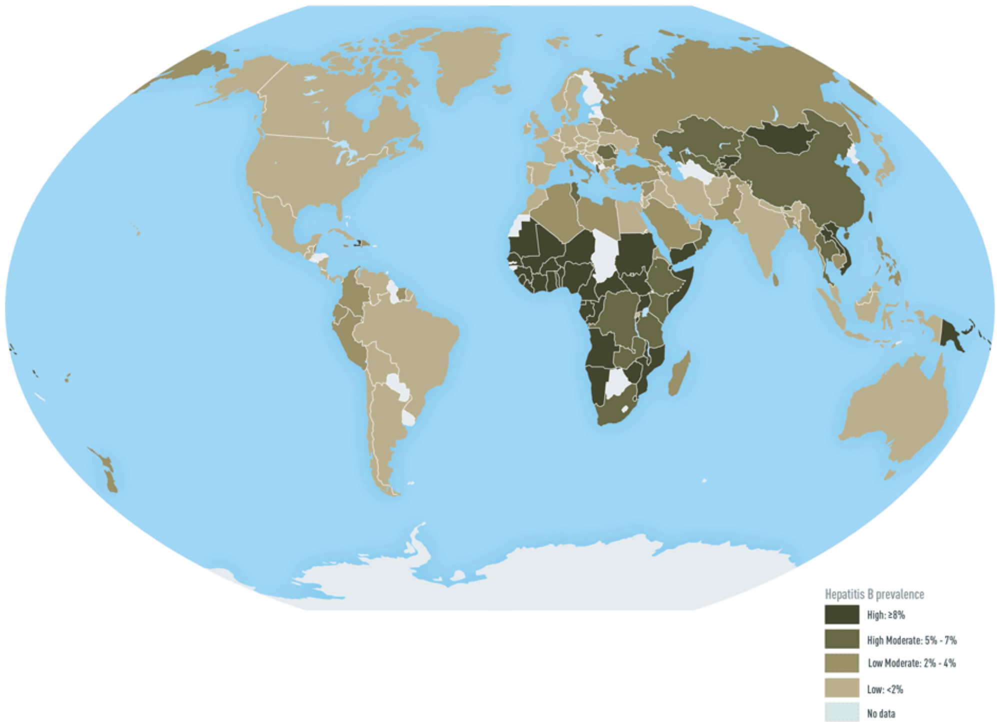
These diseases that are transmitted from animals to humans, or, from one animal to another represent a significant source of morbidity and mortality, not only for humans, but also for animals, which makes also a public health that concerns other areas of public health and medicine such as it is veterinary medicine. It has been mentioned also, that new epidemics in the world would be linked to diseases in animals or, zoonotic diseases. [12]

**The concept of carriers**

**Carrier**, is the process that can contribute to the spread of disease. A carrier contains, spreads, or harbors an infectious organism. It is accepted that **there are at least six types of carriers:** active, convalescent, healthy, incubatory, intermittent and passive. Some of them will be defined as follows:

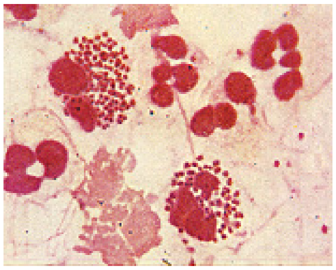
**Active carrier**, this term refers to an individual who has been exposed to and harbors a disease-causing organism and has done so for some time, even though they may have recovered from the disease. An example of this is that, worldwide as 2020, there is an estimated 257 million chronic carriers of Hepatitis B virus.

12. Myrna E. Watanabe, M.E. (2008). Animal Reservoirs: Harboring the Next Pandemic, BioScience, 58, (8), 680–684. From https://doi.org/10.1641/B580803




‘Prevalence of Hepatitits B virus worldwide,’ image from CDC.

**Convalescent carrier**, in this case, the term refers to an individual who has been exposed to and harbors a disease-causing organism (pathogen), it is in the recovery phase of the course of the disease but is still infectious. An example of this is a person recently diagnosed with COVID-19 who is in the recovery phase with no symptoms but for whom the blood test shows that the person is still infectious. **Healthy carrier**, the situation refers to an individual who has been exposed to and harbors a disease-causing organism (pathogen) but has not become ill or has not shown any of the symptoms of the disease. An example of this is, the prevalence of healthy carriers of N. meningitidis in a unvaccinated population with high serum titles of the pathogen but asymptomatic.



"An electronic picture (microscopic) of N. Menigiditis, ' Image from Blogspot.

**Incubatory carrier**, this happens with an individual who has been exposed to and harbors a disease-causing organism (pathogen), is in the beginning stages of the disease, is showing symptoms, and has the ability to transmit the disease. While looking at this and the rest of the definitions is probably useful to look at the concepts of incubation period and related terms discussed later in the chapter.

**Intermittent carrier,** this is when an individual or, animal; who has been exposed to and harbors a disease-causing organism (pathogen), and who can intermittently spread the disease at different places or intervals. Examples, chronic salmonellosis (S. typhi/enterica) in human and animals (reptiles, exotic pets, and cattle). The human or animal shed the salmonella bacteria in their feces, and if the persons with this problems don't wash its hands after defecation, the disease can be spread to others, as it is the case, of cooks who cook and serve food for others. For this reason, it is a public health measure to regularly check people working in the food industry, especially cooks for salmonellosis, so, the intermittent or chronic carrier does not infect passively other people.[13]

Figure thumbnail gr1

'Geographical distribution of typhoid fever,' Image published in Lancet, but originally developed by the World Health Organization, Public Domain.

The last of these is, the **Passive carrier**, an individual who has been exposed to and harbors a disease-causing organism (pathogen), but has no signs or symptoms of the disease, which is the same as a *healthy carrier*. A person infected with salmonella typhi can be also a passive carrier, and not only an intermittent carrier as it has been mentioned above.

### How an infectious agent entries in the body

Other information that is commonly covered in a principles of epidemiology

13. Gunn, J. S., Marshall, J. M., Baker, S., Dongol, S., Charles, R. C., & Ryan, E. T. (2014). Salmonella chronic carriage: epidemiology, diagnosis, and gallbladder persistence. Trends in microbiology, 22(11), 648–655. https://doi.org/10.1016/j.tim.2014.06.007

textbook is the concepts of how disease entry to the body or, organism and how it lefts (or, exit) the host; this is known as, **Portals of Exit and Entry** of infectious diseases. Most books and similar offer complex tables of information that can be found elsewhere,[14],[15],[16] but the following image should summarize what most student need to know for a principles of epidemiology textbook. I developed the following slide using a mnemonic that will assist with remembering the main portals, the mnemonic is COCONUT.

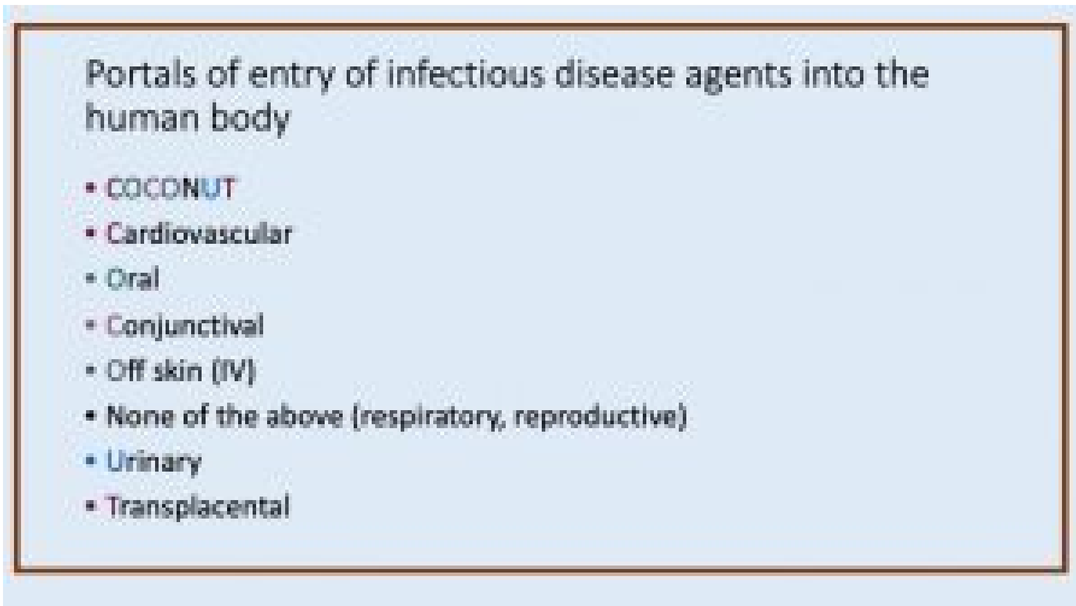



Image prepared by Giovanni Antunez, Licensed [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

Some of the entrance for disease in an organism, or, host are more effective than others, for example, blood (Intravenous, or, trans-placental) is probably the most efficient, followed by inhalation and oral.

**There are other definitions (or, terms)** that are related to the concept of disease transmission, the most important will be reviewed. These are, incubation period, inapparent infection, generation time, herd immunity.

14. BrainKart. (n.d.). Mode of Transmission In Chapter: 12th Nursing : Communicable Diseases. From https://www.brainkart.com/article/Mode-of-Transmission_37876/
15. Al Sayyari, A. (n.d.). Patient Care In Medical Imaging RAD 233. Power Point presentation available at https://slideplayer.com/slide/6400979/
16. No author. (n.d.). 3. Chain of Infection, page 3, PDF document available at https://www.gov.nu.ca/sites/default/files/files/3_%20%20Chain%20of%20Infection%20-%20march%205%20-%20low%20res.pdf

**Incubation period**, most (if not all) infectious diseases have a period in which the signs and symptoms are not manifested. The person has the infection but it is asymptomatic. Most of the time, this period called, incubation period helps the clinician and the epidemiologist to prepare in advance for the appropriate response to a disease outbreak. The incubation periods for common infectious diseases are also presented in several tables that can be found online, or, in epidemiology textbooks, for now, let's use a slide that I prepared for this topic.

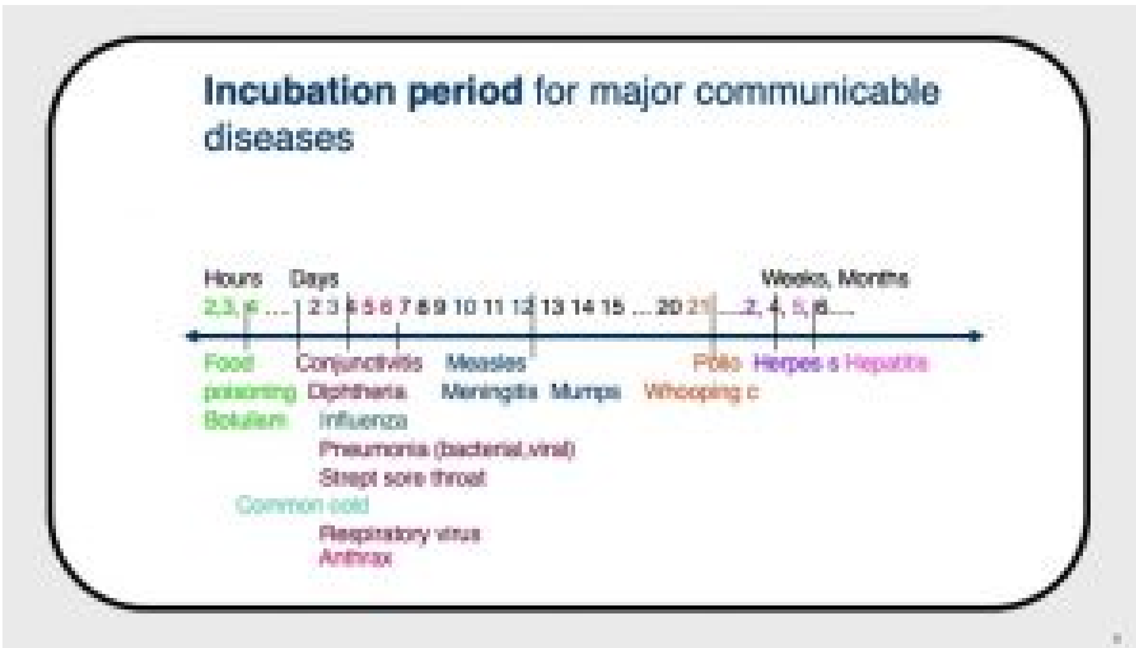



Image prepared by Giovanni Antunez, Licensed [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

As seen in the image, some infectious diseases have such as short incubation period while for others, it may take moths, and years to manifest. In the case of food contamination (commonly called, food poisoning), the incubation period is a matter of hours, and depending of the infectious agent, the severity of the disease may lead to dehydration, complications, or even death. Another example of short incubation period is the common cold, the person exposed (and contaminated) to the infection may start the symptoms as early as 12-24 hours after exposure.

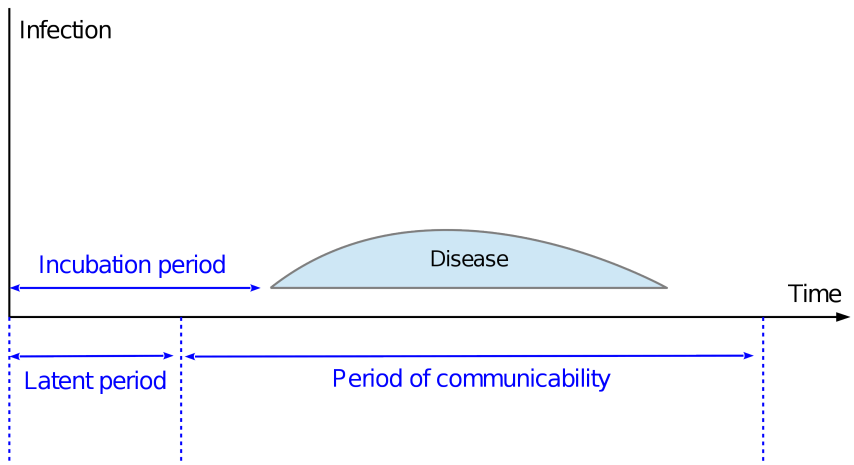
**Inapparent infection**, this is a concept that has clinical significance, because it refers to asymptomatic persons who have the disease but its condition has not reached the level to be clinically obvious (with symptoms). With the advent of COVID-19, this category has raised attention because asymptomatic individuals can transmit the disease to other susceptible hosts. These individuals with no symptoms can be unknown **carriers**.[17]

**Generation Time**

In general, this term refers to the time that takes an infectious disease to elevate at the level of **case**, which is usually when the full set of symptoms manifest, and the person seek health care. Another way to see the term, is to look at the time interval between the presence of an infectious agent in a host and maximal time of communicability. The practical application of this concept is that for some diseases the maximum period of communicable can be longer, while for others is short. Both assists the epidemiologist or, any other public health professional to work in the prevention and control of infectious diseases.

Another way to see the process described above is presented in the image below:

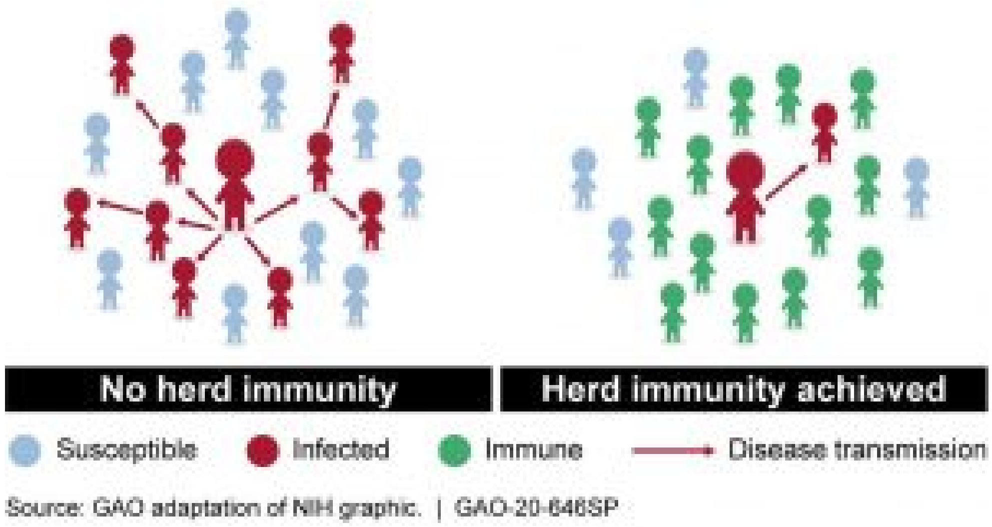
17. Syangtan G., Bista S., Dawadi P., Rayamajhee B., Shrestha LB., Tuladhar R., Joshi DR. (2021). Asymptomatic SARS-CoV-2 Carriers: A Systematic Review and Meta-Analysis. Frontiers in Public Health, 8, 1066. From https://www.frontiersin.org/article/10.3389/fpubh.2020.587374




‘Concept of Incubation period,’ image from Wikimedia, Licensed CCO 1.0 Public Domain.

**Herd immunity**

In general and for those of us whose English is a second language, the word ‘herd’ can be confusing, but one way to understand its usage is when we refer to groups of sheep, cattle or, any other ‘herd’ that needs protection (from the wolf, or, the possum). So, the word **'herd'** is mostly synonymous of protection. In simple terms, herd immunity refers to the percentage of people or, animals that can be protected by immunization. The higher the number of immunized individuals, the higher the protection for those who are not immunized. In the past, this concept was highly regarded, and it was the belief that with at least 70% of vaccinated individuals, the protection was accomplished. But over the years, and with the advent of so many infectious diseases (old, and new), there is some consensus that the number should be at least 95% (the number varies by disease) of vaccinated in order to get herd immunity for those who are not immunized.[18], [19], [20], [21]




'A graphic representation of the concept of herd immunity', This image is excerpted from a U.S. GAO report.

It is important to mention that herd immunity does not work for all infectious diseases, that is another reason for the concept to be considered a 'weak' and 'old' concept to refer to this method (vaccination) used to control the spread of an infectious disease. Maybe the key is to add to the vaccination other effective features that have been used in public health for centuries, good sanitation, soap and water to wash hands and other surfaces, mask covering and social distance among other measures that we have seen working during the COVID-19 pandemic.[22].

18. MacMillan, C. (May 21, 2021. Herd Immunity: Will We Ever Get There? From https://www.yalemedicine.org/news/herd-immunity
19. The Vaccine Knowledge Project. (n.d.). Herd immunity (Herd protection). From https://vk.ovg.ox.ac.uk/vk/herd-immunity
20. Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R., & Bettinger, J. (2013). Vaccine hesitancy: an overview. Human vaccines & immunotherapeutics, 9(8), 1763–1773. From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3906279/
21. Association for Professionals in Infection Control and Epidemiology (APIC). (2021). Herd immunity. From https://apic.org/monthly_alerts/herd-immunity/
22. D'Souza G., Dowdy, D. (September 13, 2021). Rethinking Herd Immunity and the Covid-19 Response End Game. Johns Hopkins University. From https://publichealth.jhu.edu/2021/what-is-herd-immunity-and-how-can-we-achieve-it-with-covid-19

**Methods commonly used to prevent and control infectious diseases (also, called, communicable diseases)**

The following section will cover the essential methods used commonly to prevent an infectious disease, it starts with the use of vaccines, sanitation, isolation, quarantine and additional measures that had been used for centuries but that are becoming relevant these days with the COVID19 pandemic. In addition, **there are three key factors** that are considered essential, 1) Remove, eliminate, or contain the cause or source of infection, 2) Disrupt and block the **chain** of disease transmission, and 3) Protect the susceptible population against infection and disease. In this context, and additional set of recommendations are also prescribed, and discussed in the following paragraphs:

**Vaccines (also, called, Immunizations)**

In the history of medicine and public health (including, epidemiology), vaccines have been considered the major weapons of defense especially on the prevention of infectious diseases, especially those common childhood infectious diseases.[23] See below a list of common vaccines that are available:

23. CDC. (n.d.). Vaccines for Your Children, Diseases & the Vaccines that Prevent Them. From https://www.cdc.gov/vaccines/parents/diseases/index.html

## Most common vaccines

| | |
|---|---|
| Anthrax | Pneumonia |
| Cholera | Polio |
| Diphteria | Rabies |
| Rubella | Smallpox |
| Hepatitis A | Spotted fever |
| Hepatitis B | Tetanus |
| Influenza | Tuberculosis |
| Measles | Typhoid fever |
| Meningitis | Typhus |
| Mumps | Whooping cough |
| Pertussis | Yellow fever |
| Plague | Chickenpox |

Image prepared by Giovanni Antunez, Licensed CC BY 4.0

In general, vaccines are used to prepare the body to resist infection. Most of them are inactivated bacteria, viruses, or microbe toxins. This process is called, an 'antigen-antibody reaction'. In other word, the vaccine acts as the antigen (a substance capable of stimulating the formation of antibodies). These antibodies protect the person against the infectious agents that are the cause of the disease. [24] See image of this process:

24. World Health Organization. (2020). How do vaccines work? From https://www.who.int/news-room/feature-stories/detail/how-do-vaccines-work

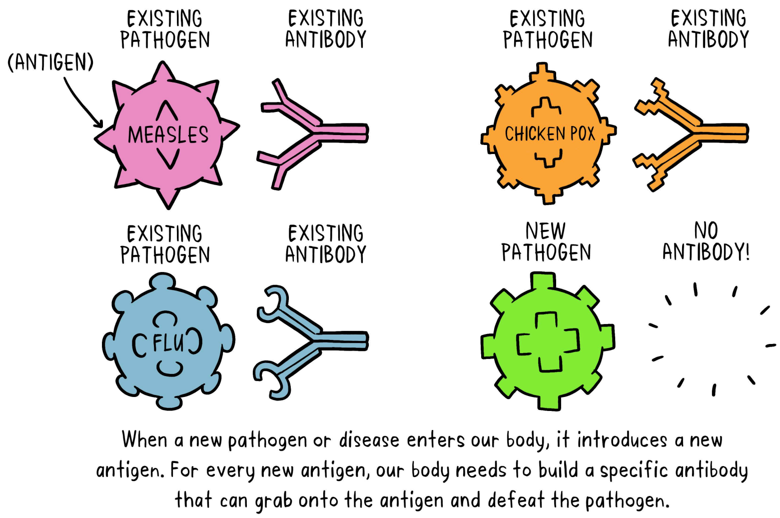



Image from how vaccines work.

Until recent years, most of the vaccines contained the infectious agent, but in recent years, especially during the COVID-19 pandemic in which two vaccines (Pfizer–BioNTech and Moderna) are developed using using mRNA technology, which changes the paradigm that a vaccine to combat this disease should contain an attenuated form of the infection agent. However, more time will be needed to use the mRNA technology to develop effective vaccines in the future, especially for new emerging, and re-emerging infectious diseases.[25]

### Sanitation

Hand washing has been the oldest of the measures used to prevent infection. For centuries, public health has promoted hand washing as the number one activity to

25. Excler, JL., Saville, M., Berkley, S. et al. (2021). Vaccine development for emerging infectious diseases. Nat Med 27, 591–600. From https://www.nature.com/articles/s41591-021-01301-0

keep people out of disease (infectious disease).[26], [27] Together with this measure, it is important to put attention to personal hygiene (frequent bathing, regular grooming, teeth cleaning, and maintenance, changing clothes frequently), and also to keep ventilation in homes, and buildings as the cleaning of surfaces that can act as **Fomite(s)** or, inanimate transmitters of an infectious agent.[28]

'Hand washing' image from Pixbay.

In addition, face mask covering and personal protective equipment or PPE are also considered

26. CDC. (no date). Show Me the Science - Why Wash Your Hands? From https://www.cdc.gov/handwashing/why-handwashing.html
27. Goldust, M., Abdelmaksoud, A., & Navarini, A. A. (2020). Hand disinfection in the combat against COVID-19. Journal of the European Academy of Dermatology and Venereology : JEADV, 34(9), e454–e455. https://doi.org/10.1111/jdv.16574
28. CDC. (2022). How to Protect Yourself & Others. From https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html

especially for health care workers who are considered more at risk due to their exposures with infected clients, and surfaces.[29]

[Family mask](), image from [Pixabay.]()

**Environmental controls to prevent infectious diseases**

This measure is aimed at providing clean and safe air, water, milk, and food. It also includes the management of solid waste (trash, and garbage); liquid waste (sewage); and control of **vectors** (insects and rodents) of disease.

**Host-related control & prevention**

Based on the concept model of the **epidemiological triangle of infection**, the host-related control is intended to protect the host from contagious diseases and infections the following protective measures are used: **quarantine** and isolation.

**Quarantine**

The original meaning of this word is that quarantine means, forty days, however in recent years, the word quarantine refers to specific amount of time considered in which a person is isolated, or, separated from those who are not infected with the agent, and it is not necessarily forty days, it can be as short a week, two weeks, depending on the situation as it has been used during the COVID-19 pandemic.[30], [31]

29. Tabatabaeizadeh S. A. (2021). Airborne transmission of COVID-19 and the role of face mask to prevent it: a systematic review and meta-analysis. European journal of medical research, 26(1), 1. https://doi.org/10.1186/s40001-020-00475-6
30. CDC. (no date). History of Quarantine. From https://www.cdc.gov/quarantine/historyquarantine.html
31. Kiliç, R., Ataman Hatipoğlu, Ç., & Güneş, C. (2020). Quarantine and its legal dimension. Turkish journal of medical sciences, 50(SI-1), 544–548. https://doi.org/10.3906/sag-2004-153

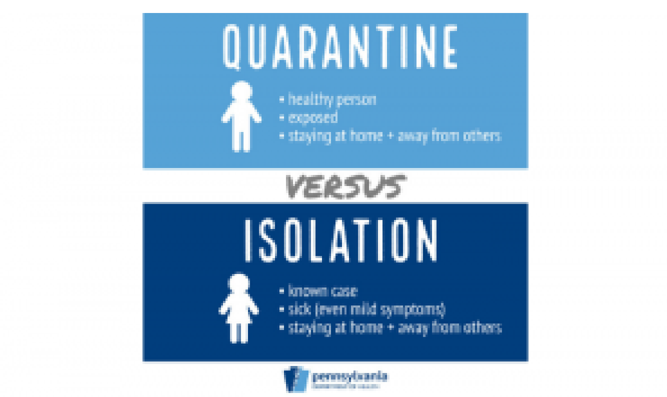
In addition, it accepted that there are four levels of quarantine used in public health: 1) Segregation, 2) Personal surveillance, 3) Modified quarantine, 4) Complete quarantine. Most of the mentioned levels are very much self explanatory but more details can be found in the Internet and epidemiology and public health literature.[32]

Quarantine PNG Free Download

'Quarantine', image from PNGmart.

**Isolation**

This method is mainly used for limited number of cases as it is in the case of humans, and also for animals. There are six levels of isolation: 1) In a private isolation room, 2) The use of separate and infection control gowns, 3) Staff must wear masks, 4) All staff must gloved with interacting, treating, or working with or on the patient or subject (e.g. an animal), 5) Hands washing is required upon entering, and leaving the patient's room, 6) All contaminated articles or possible contaminated articles including linen, dressing, syringes, instruments, etc., must be disposed properly.[33]




'Quarantine versus isolation', Image from Dept. of Health, City of Philadelphia, Pennsylvania.

A combination of quarantine and isolation is commonly seen especially in clinics and hospitals, not only during the COVID-19 pandemic but also during the management of other public health emergencies dealing with highly contagious diseases such as Cholera, Ebola, Marburg virus, severe cases of tuberculosis, etc.[34]

32. CDC. (no date). Quarantine and Isolation. From https://www.cdc.gov/quarantine/index.html
33. CDC. (No date). Isolation. From https://www.cdc.gov/coronavirus/2019-ncov/your-health/quarantine-isolation.html#iso
34. Barbisch, D., Koenig, K. L., & Shih, F. Y. (2015). Is There a Case for Quarantine? Perspectives from SARS to Ebola. Disaster medicine and public health preparedness, 9(5), 547–553. https://doi.org/10.1017/dmp.2015.38

In the practice of epidemiology, all of the concepts discussed in this chapter are used for disease surveillance and disease investigation, especially disease outbreaks in which very common to be assisted by the use of the concept and calculations known as attack rate, which is discussed in detail in the final chapter of this textbook.

**Summary**

This chapter has covered the most essential concepts and definitions commonly used in the area of infectious diseases epidemiology. As it has been repeated several times in the content above, the study of infectious diseases is the original focus of the field of epidemiology, which is the main reason for having a separate chapter on the topic. The medical model is the dominant force behind the study of infectious diseases, and the majority of the methods used in prevention and control of these diseases. Infectious diseases constitute the major focus of the public health/health care system activities and interventions in the United States and probably internationally.

# 4. Person, Place and Time

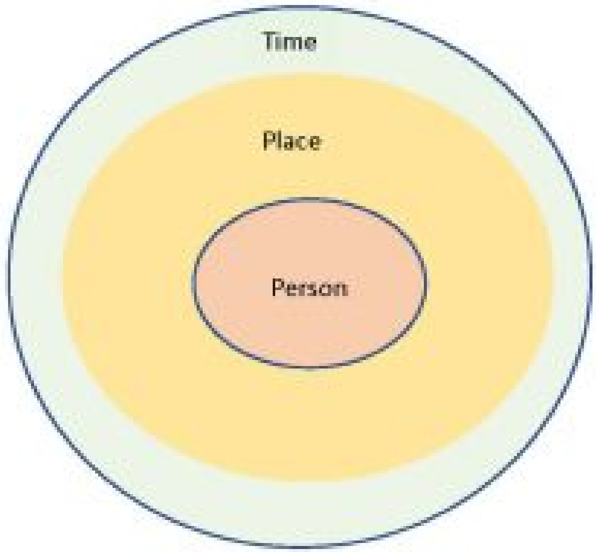
*Learning Objectives*

By the end of this chapter, the learner will be able to

- Learn about the major subdivision of epidemiology, descriptive and analytic.
- Describe the importance of person, place, and time as the main variables used in descriptive epidemiology
- Set the foundations for the understanding of new concepts that will be study later especially the field of analytic epidemiology.
- Appraise the importance of the concepts of person, place and time as fundamental for the understanding of health phenomena and its related outcomes.

**Person, Place and Time**

These three are considered the most common variables used in the study of epidemiology. Researchers in epidemiology and public health use these three variables (person, place and time) to look for associations and health determinants that explain several health phenomena, and especially illness. Frequently, it is important to know who the **person** is – although, it can be also an animal, insect, or, any other living, or, non living thing who has been involved in a specific event, or, incident; and the **place**, and **time** in which this had occurred. It is kind of similar to what journalists use in the writing/reporting of their news, the who, where, and when. Who – the person, Where, the place, and When, the time. Although I had introduced the concepts of person, place and time in a simplistic manner in the previous paragraph, the study of these three variables is not as simple as it appears, there are several sub-variables that also used to make sense of the information, and to go into more details, is the main purpose of this chapter.




Person Place and Time. Image created by Giovanni Antunez. Licensed CC BY 4.0

**Characteristics of Person**

The Who. Who is affected? or, Who is at risk? etc. Commonly the who is a person, and that is the reason for the name of it, person. But in some cases for example in veterinary medicine, the who is an animal; and in environmental health, the who could be an ecosystem, or, a lake, or, a coastal region. But most of the time, the person refers to a human. Several of the characteristics of person match the 'individual lifestyle factors' category commonly listed in an overview of the social determinants of health. One example of this is the following image:

**Common characteristics of Person**

The most common subcategories of the variable person are, race and ethnicity; gender, age, occupation, marital status, social economic status, and other socio economic, and political variables.[1] Some of these characteristics may have a major impact on the health of a person more than others, but in general, all of them are important as it is a combination of factors (also called, determinants) that 'determine' the health of the individual. For example, no one can denied that socioeconomic conditions are extremely important in the life of a person, so, it is possible than more than one problem can be explained by this variable, but if every problem could be explained by only one indicator such as socioeconomic status, then, the measurement of this variable became confusing, or, not to say, useless; for this reason it is important to keep in mind that it is a combination of factors (variables, determinants) that explain the health status of the individual, the community, and the nation.

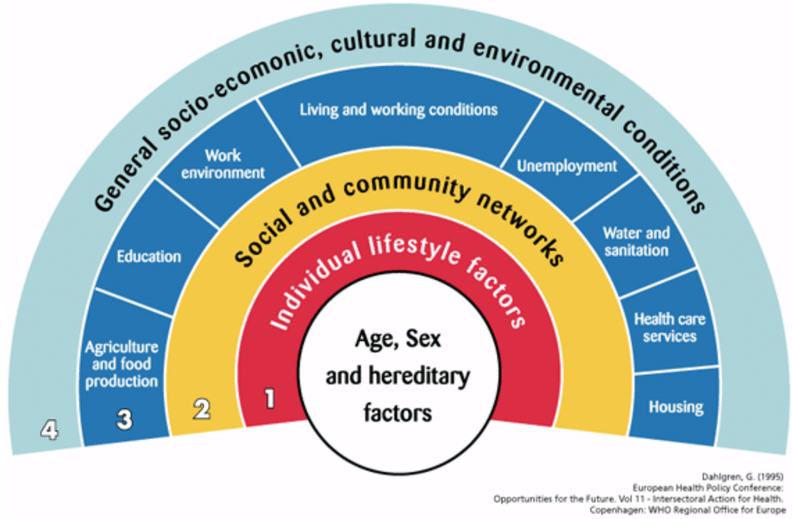



Image from Health Knowledge

**Gender**

The category, or, variable gender has been frequently misused as an equivalent of sex, but in reality sex is not the same as gender. Again, the National U.S. Census still uses the category sex to report, male, or, female; but is essentially incorrect. People are either male or, female; but being male of female has more to do with biology because there is also a wide spectrum of variations as current research reports. See for example a recent image from an online publication that clearly shows that

1. Anderson NB, Bulatao RA, Cohen B, editors. (2004.). National Research Council (US) Panel on Race, Ethnicity, and Health in Later Life, In Critical Perspectives on Racial and Ethnic Differences in Health in Late Life. Washington (DC): National Academies Press 9. From: https://www.ncbi.nlm.nih.gov/books/NBK25526/

progress have been made on this area, especially the use of personal pronouns to refer to the gender category.[2]


Image from Gender-identity

For reasons of conventionality, in epidemiology, the category gender is still referred as, sex. As an example of this category, it is recognized in public health and other fields of study that gender as a health determinant has a strong influence in the health status of women, especially in issues related to pregnancy and childbearing; it has been also reported that women tend to suffer more of series of chronic diseases that affect them differently compared to men.[3]

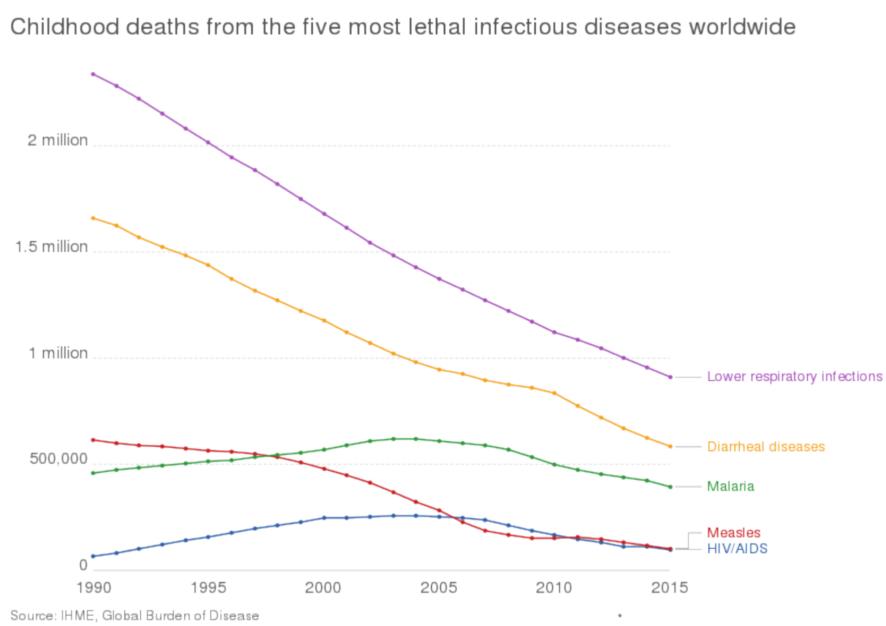
### Age

In a broad (general) manner, the variable age seems to be self-explanatory, since we know that people in general are children (with several subcategories categories), youth, adults, or, seniors.

2. AAUW. (n.d.). Dimensions of Diversity & Identity In DEI Toolkit: Gender & Gender Identity. From https://www.aauw.org/resources/member/governance-tools/dei-toolkit/dimensions-of-diversity/gender-identity/
3. Iassoff C. (2007). Gender differences in determinants and consequences of health and illness. Journal of health, population, and nutrition, 25(1), 47–61. From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3013263/

Image from [Freepik](https://www.freepik.com)

But age is a complex variable especially for data collection, and analysis purposes, that is because age in a person defines several health stages. Epidemiology is concerned with the presence of disease for people in all ages, but the childhood and adulthood period are of such as concern because common knowledge says that are diseases that are characteristic of these two age periods. For example, immune preventable diseases such as polio, chicken pox, measles, etc., are common childhood diseases, and degenerative diseases especially those related to memory (such as dementia and similar neurological disorders) are more common in old age.




"Childhood deaths from the five most lethal infectious diseases worldwide", Our World In Data, CC 3.0 Unported license.

There are several health issues that concern only to specific ages. The graph above shows the most common childhood diseases that are the major causes of childhood mortality in the world. While the next chart shows common diseases of the elderly population in the European region.

Image from From the World Health Organization (WHO)

**Occupation**

A person's occupation seems to be an obvious category because in general, people and their occupations are defined (in a broad way) by their jobs, or, work. But what about the people who work at home such as home makers, or, people who are unemployed, or, underemployed? This question raises the point that occupation is not as simple, and obvious as we generally think. If there is a place in which the variable occupation matters, is in the field of environmental sciences. Exposures to toxic chemicals, and other environmental contaminants are higher in certain occupations, or, jobs.[4]

**Occupation during pandemic times in the U.S.**

In the year 2020 affected by the COVID-19 pandemic, the rate of unemployment in the U.S. rose as a consequence of the complete disruption of people's lives, and companies that went out of business due to the restrictions imposed by public health state and national ordinances.

See image:[5]

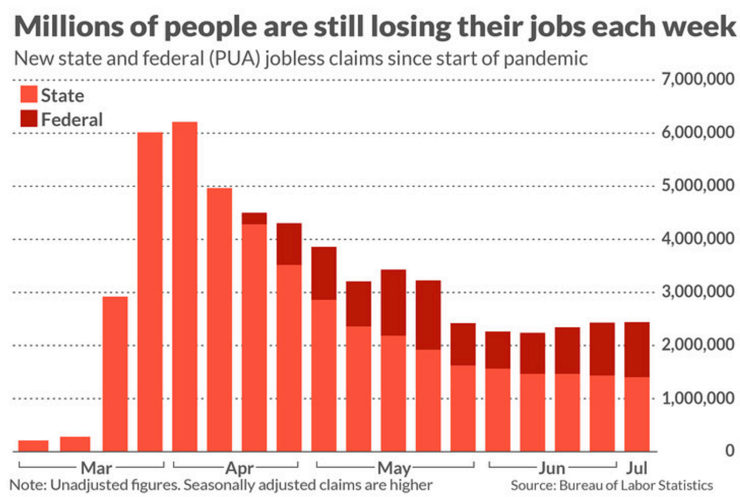
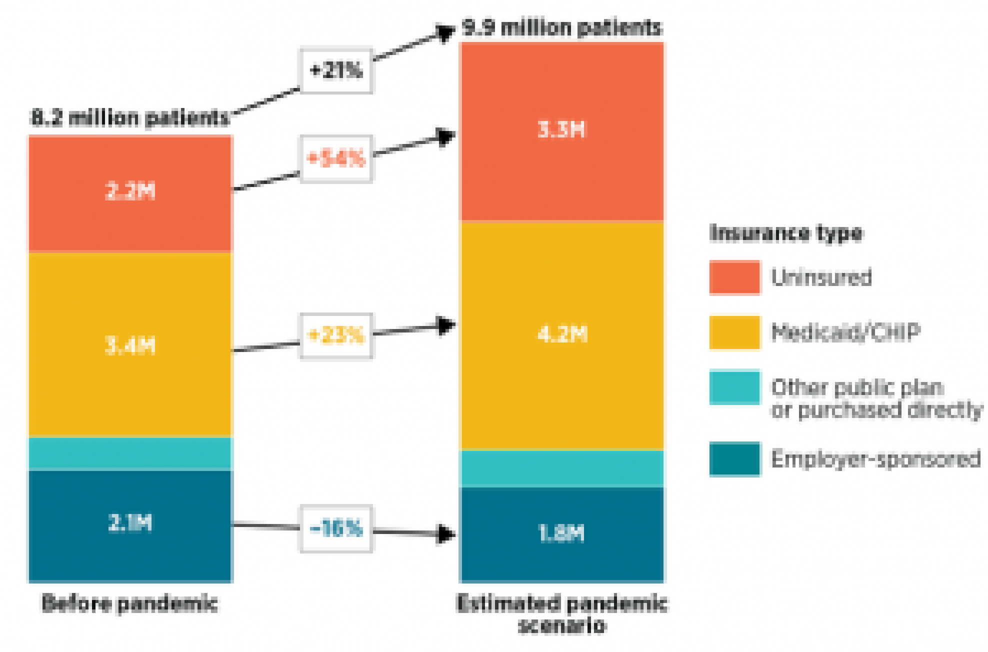



Image from the [Census Bureau](https://www.census.gov).

**Unemployment** affects the health of individuals in different manners, but one area that is highly affected is mental health, and also the access to health care services, problems in paying rent, not to say the ability of buy food, and pay for other personal or, family needs. In the case of access to health services, there a very recent example is the case of women seeking sexual reproductive health services at publicly funded health clinics,

4. Foulis, M. (October 22, 2020). 7 most common occupational diseases, what is an occupational disease? And which are the most common? Canadian Occupational Safety. From https://www.thesafetymag.com/ca/topics/occupational-hygiene/7-most-common-occupational-diseases/236947
5. Bartash, J. (July 9, 2020). Jobless claims tell us 33 million people are unemployed, but many doubt it's that bad. In Market Watch. From https://www.marketwatch.com/story/jobless-claims-tell-us-30-million-people-are-unemployed-but-many-doubt-its-that-bad-2020-07-08

according to insurance status before and during the COVID-19 pandemic. See image below:[6]




Unemployment in the United States before and during pandemic.
Image from [HealthAffairs](https://www.healthaffairs.org)

### Marital status

In general marital status more than any other variable is mainly a legal category, or, another example of social construct. What defines a person as single or, married is controversial if we think that two people living together for a long period of time are a couple, and although they may not be legally/officially married, they are in a married status. And, it is for this reason that in some states in the U.S. after 15 years of living together, for legal purposes, a couple is considered married even if they are not officially married at the beginning of their relationship. For practical (operational) reasons in epidemiology, the marital status categories used in research are the U.S. National Census categories, which are four major categories: never married, married,

6. Sonfield, A., Frost, JJ., Dawson, R., Lindberg LD. (August 3, 2020). COVID-19 Job Losses Threaten Insurance Coverage and Access to Reproductive Health Care For Millions In Health Affairs Forefront. From https://www.healthaffairs.org/do/10.1377/forefront.20200728.779022/full/

widowed, and divorced.[7] Details of these mentioned categories are summarized in the table that follows:

7. U.S. Census Bureau. (n.d.). Subject Definitions, Marital Status. From https://www.census.gov/programs-surveys/cps/technical-documentation/subject-definitions.html

| Category | Descriptor |
|---|---|
| “Married” | Married category is divided into “married, spouse present,” “separated,” and “other married, spouse absent.” A person was classified as “married, spouse present” if the husband or wife was reported as a member of the household, even though he or she may have been temporarily absent on business or on vacation, visiting, in a hospital, etc., at the time of the enumeration [during the census data collection.]. |
| “Separated” | People reported as separated included those with legal separations, those living apart with intentions of obtaining a divorce, and other people permanently or temporarily separated because of marital discord. |
| “Other married” | The group “other married, spouse absent” includes married people living apart because either the husband or wife was employed and living at a considerable distance from home, was serving away from home in the Armed Forces, had moved to another area, or had a different place of residence for any other reason except separation as defined above. |
| “Single” | Single, when used as a marital status category, is the sum of never-married, widowed, and divorced people. “Single,” when used in the context of “single-parent family/household,” means only one parent is present in the home. The parent may be never-married, widowed, divorced, or married, spouse absent. |

**Table of Marital Status categories.** Content prepared by the author using the U.S. Census Bureau marital status categories.[8]

### Race and Ethnicity

Although it is accepted that race is a social construct. It has an impact on the life of people more than any other, and it is one of the major factors used to explain health disparities especially in the United States. The concepts of race and ethnicity had developed so much in recent years in this country. Theoretically, it has been known for decades that race and ethnicity are complementary categories, but not the same. Race is mostly biological characteristics of the person, while ethnicity relates mostly to the ethnic group from which the person descend; both are socially constructed concepts. As the American Sociological Association defines, "Race" refers to physical differences that groups and cultures consider socially significant, while "ethnicity" refers to shared culture, such as language, ancestry, practices, and beliefs.[9] In this class, the classification used for epidemiological concepts, terms, statistical reasons (including categorization of ethnic groups in the U.S.), and calculations is, the Office Bureau of Census race classification[10] which essentially assign the U.S. population in the following five categories:

*Key Takeaways*

The standards have five categories for data on race: **American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White**.

Although updates have been made in the 2021 census report, in general, these categories have been defined since 1997 [11], these basic definitions follow:

8. U.S. Census Bureau. (n.d.). Marital Status. From https://www.census.gov/programs-surveys/cps/technical-documentation/subject-definitions.html#maritalstatus
9. American Sociological Association. (n.d.). Race and Ethnicity. From https://www.asanet.org/topics/race-and-ethnicity
10. Office Bureau of Census. (n.d.). Race Categories. From https://www.census.gov/
11. U.S. Dept. of the Interior, Office of Civil Rights. (n.d.) Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity. From <a href="https://www.doi.gov/pmb/eeo/directives/race-data

**This table has been prepared using the U.S. Bureau of Census[12] five categories used for data reporting on race and ethnicity.**

| Description | Category |
|---|---|
| A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment. | ***American Indian or Alaska Native*** |
| A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. | ***Asian.*** |
| A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American." | ***Black or African American*** |
| A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term, "Spanish origin," can be used in addition to "Hispanic or Latino." | ***Hispanic or Latino*** |
| A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. | ***Native Hawaiian or Other Pacific Islander*** |
| A person having origins in any of the original peoples of Europe, the Middle East, or North Africa. | ***White*** |

12. U.S. Dept. of the Interior, Office of Civil Rights. (n.d.) Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity. From https://www.doi.gov/pmb/eeo/directives/race-data

In 2010, there was change/addition to the five categories by creating the new, **"Some Other Race"** category, which includes all of the other responses (when a person responds to the Census form) not included in the five categories listed above (in the table). The person can respond with entries such as being, multiracial, mixed, interracial, or, Hispanic or belonging to a specific Lation group such as Mexican, Puerto Rican, Cuban, or, Spanish).[13]

### From the 2010 national census to now

From 2010 to 2020 lots have changed in the U.S. and the 2020 U.S. national census reflect these changes, and acknowledged than the country has become more diverse than ever. To arrive to this conclusion, the Census Bureau used what they called, the 'Diversity Index (DI),' not discussed in detail here in this book, but this index should reflect the changes experienced by the country in recent years, and especially in the last decade.[14]

### Religion

Religion in the health sciences and especially in public health is considered a protective factor[15], which means that religion (as a factor, or, state/condition) help the individual and community to decrease the risk of development negative health outcomes. For example studies show that religion is a protective factor against drug use,[16] It is important to note here, that the article cited here, refer to 'religiosity,' and not to religion in general, which is an important distinction. Because it is not the religion itself but the lifestyle associated with the religious practice. In this context, '**religion**' is considered a social determinant of health, but since the question, is it causal? More than causal, it is more of what is called, **reversed causation** and a series

13. Parker, K., Horowitz, JM., Morin, R., Lopez, MH. (n.d.). Chapter 1: Race and Multiracial Americans in the U.S. Census. Pew Research Center. From https://www.pewresearch.org/social-trends/2015/06/11/chapter-1-race-and-multiracial-americans-in-the-u-s-census/
14. Office Bureau of Census. (n.d.). Racial and Ethnic Diversity in the United States: 2010 Census and 2020 Census. From https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html
15. No author. (n.d.). Protective Factor. From https://en.wikipedia.org/wiki/Protective_factor
16. Van der Meer Sanchez Z, De Oliveira LG, Nappo SA. (2008). Religiosity as a protective factor against the use of drugs. Subst Use Misuse, 43(10):1476-86. From https://pubmed.ncbi.nlm.nih.gov/18615320/

of confounding factors. Very limited studies (especially clinical trials) have been conducted to assess religion and its positive effect on the health of an individual. [17].

In a sense, it can be said that more than religion [as an entity itself], it is 'spirituality' that plays a role in the health of the individual, as an example taken from the field of psychology shows, see the visual below – there is a direct relationship between spirituality and health-related behaviors, and the same can be said, between spirituality and psychological well-being.[18]

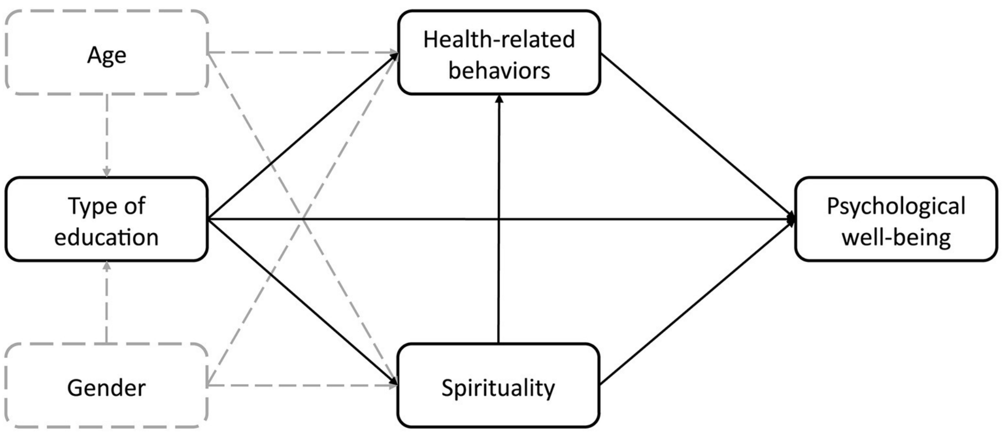



Image from [Frontiers in Psychology](https://www.frontiersin.org/article/10.3389/fpsyg.2020.01997).

### Socio-economic status

Socio economic status is related to social class. In general, groups in society that have medium to high levels of income tend to have a better place in society compared to those who don't, but from the public health perspective, the importance of socioeconomic status is a predictive factor of the lack of resources to stay healthy, access to health care services, and also access to healthy foods just to mention some of the essentials for a healthy life.[19]

17. Kawachi, Ichiro. (2020). Invited Commentary: Religion as a Social Determinant of Health, American Journal of Epidemiology, 189 (12), 1461–1463. From https://doi.org/10.1093/aje/kwz204
18. Bożek, A., Paweł FN., Blukacz M.(2020). The Relationship Between Spirituality, Health-Related Behavior, and Psychological Well-Being. Frontiers in Psychology, 11, 117. From https://www.frontiersin.org/article/10.3389/fpsyg.2020.01997
19. Azétsop, J., & Joy, T. R. (2013). Access to nutritious food, socioeconomic individualism and public health ethics in

**Socioeconomic status Categories**

For reasons of **operationalization**, this variable is usually divided into three levels (**high, middle, and low**), in this form researchers, and the general public situate individuals, families and communities (in the broad sense) in relation to others.[20]

On the other hand, for the mentioned levels of **income** (earnings), the Bureau of Census assigns a specific value using what is called, the **Gini ratio** (or index of income concentration) is a statistical measure of income equality ranging from 0 to 1. A measure of 1 indicates perfect inequality; i.e., one person has all the income and rest have none. A measure of 0 indicates perfect equality; i.e., all people have equal shares of income. The Census Bureau used grouped data to compute all Gini ratios. For a more detailed discussion, see Current Population Reports, Series P-60, No. 123.[21].

Since inequalities are found, the following graph (taken from the U.S. Census Bureau) reflects the income by gender in the U.S. in 2020. [22]

the USA: a common good approach. Philosophy, ethics, and humanities in medicine : PEHM, 8, 16. From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4231366/

20. Worthy, L.D., Lavigne, T., Romero, F. (2020). Socioeconomic status In Culture and Psychology, Maricopa Community College, Phoenix, Arizona. From https://open.maricopa.edu/culturepsychology/chapter/socioeconomic-status-ses/
21. U.S. Census Bureau. (n.d.). Income Measurement. From https://www.census.gov/programs-surveys/cps/technical-documentation/subject-definitions.html#incomemeasurement
22. U.S. Census Bureau. (2021). Consumer Income (P60) Publication Series. Figure 4 in the respective report. From https://www.census.gov/content/dam/Census/library/visualizations/2021/demo/p60-273/figure4.pdf

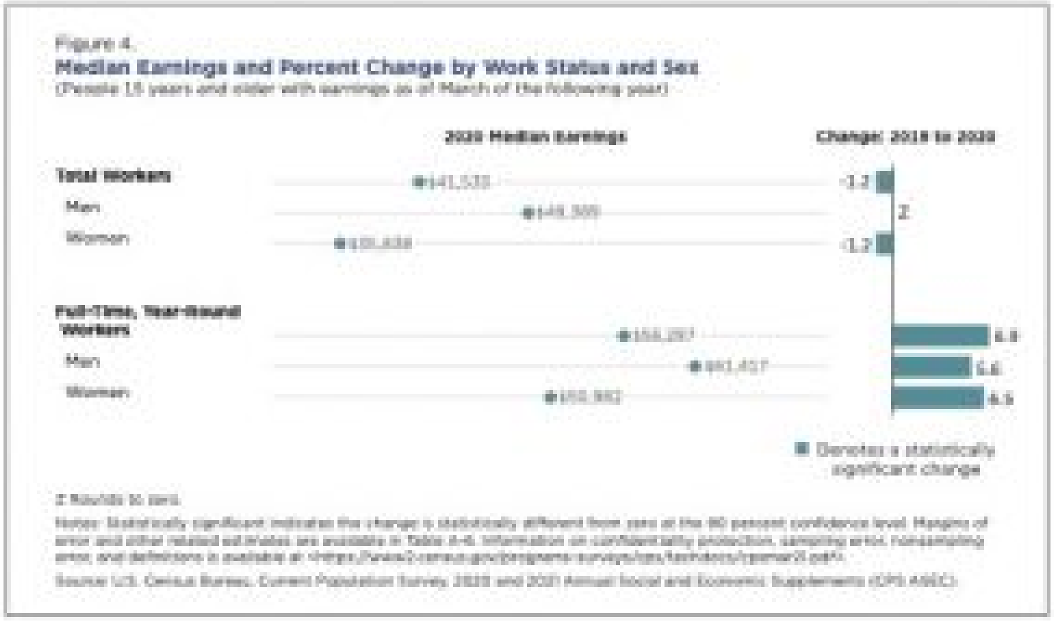


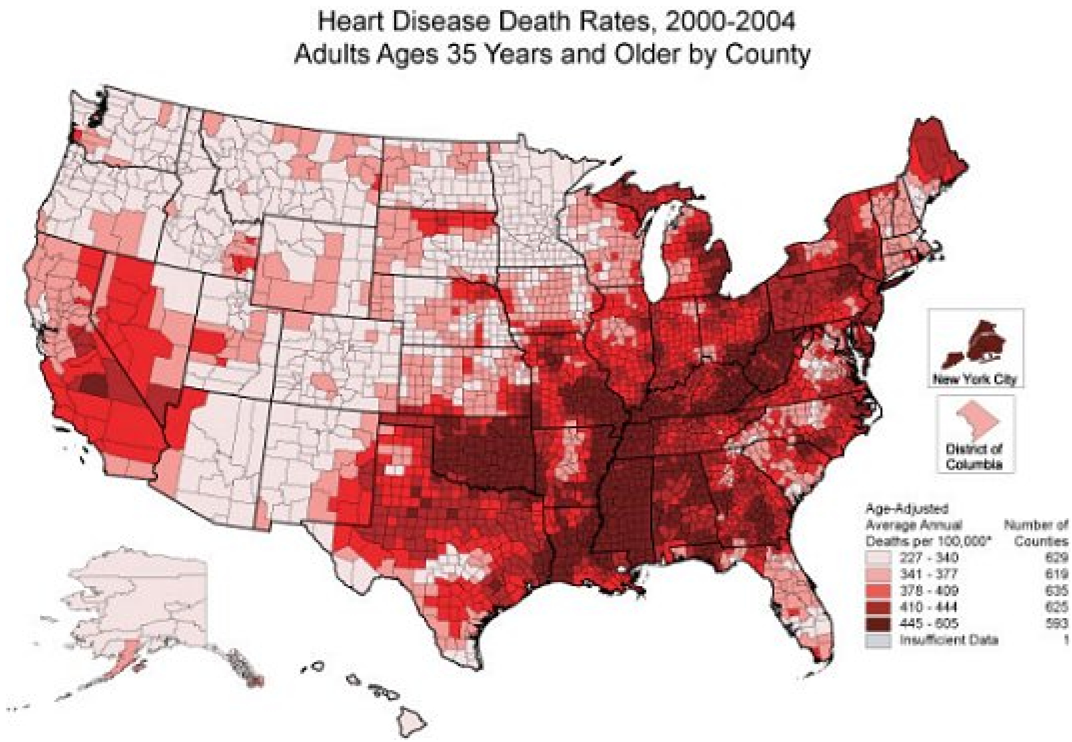

Image from the [U.S. Census Bureau.](#)

## Characteristics of Place

As it was said at the beginning of this chapter, in epidemiology, the use of 'Who?,' 'Where?,' and 'When?' helps to remember that the 'Who? is the person, the 'Where' refers to the place, and 'When?' The time.[23] In this section, the focus is on the 'place' or, the Where?

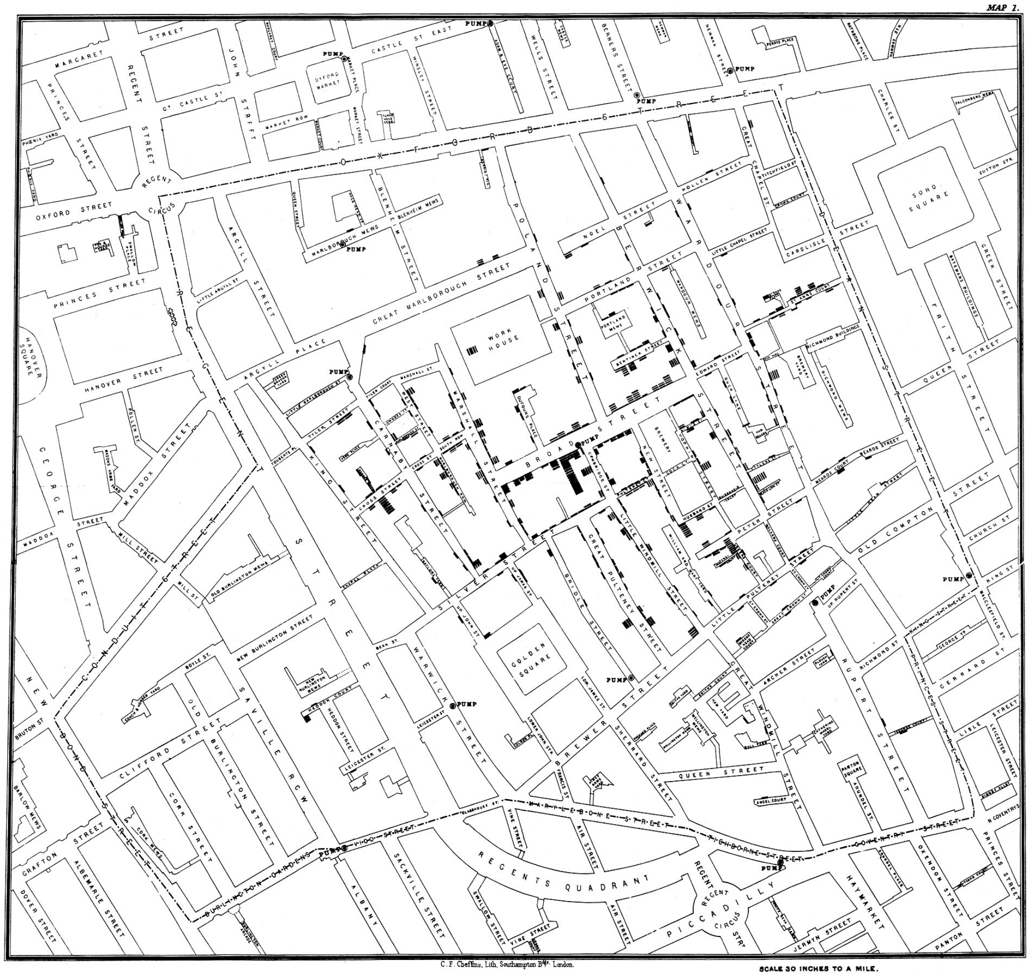
One of the major characteristics of place is that it could mean more than one thing, for example, place can refer to a location (an area, a city, a state, a country, etc.), but since place is a spatial concept, it is frequently described using geospatial coordinates such as latitude and longitude. That is one of the main reasons, in recent decades, geospatial analysis have grown so much in the field of epidemiology and public health in general. [24]. As an example, map of heart disease in the U.S. that was developed using **GIS** (or, geographic information system) software is presented below:

23. Centers for Disease Control and Prevention. (n.d.). Lesson 1: Introduction to Epidemiology Section 6: Descriptive Epidemiology in Principles of Epidemiology for Public Health Practice. From https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section6.html
24. Jacquez, G. (2000). Spatial analysis in epidemiology: Nascent science or a failure of GIS?. J Geograph Syst 2, 91–97. From http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.468.1483&rep=rep1&type=pdf




Heart Disease Death Rates in the U.S. Image from CDC, Public Domain.

Since the image is developed using data that is represented by dots (the pixels in the image), another name for this type of image is, **dot map**, which was originally introduced (and probably created for the first time) by Dr. John Snow 1854 cholera epidemic, he draw the map without using any scale, because all he wanted was to designated where the major sources of infection where coming from in the city of London. See this map below:




[Dr. Snow's map showing cholera cases in London during the epidemic of 1854.]() Image from [Wikipedia](), [Public Domain]().

## Common characteristics of place

### Urban/Rural

The definitions of urban and rural creates some confusion, what is rural? And when rural ends and urban begans? This has been a growing discussion in recent years since most urban areas in the United States have grown so much that rural areas have been engulfed by the metro area. The U.S. Bureau of Census classify communities as urban and rural by relating them to Metropolitan Statistical Areas (MSAs) and census tracts.

In this context, it is customary in epidemiology and public health to use the NCHS urban-rural classic classification that uses six levels (of categories) as follow: [25]

25. National Center for Health Statistics (NCHS). (n.d.). NCHS Urban-Rural Classification Scheme for Counties. From https://www.cdc.gov/nchs/data_access/urban_rural.htm#2013_Urban-Rural_Classification_Scheme_for_Counties

| Category | Description |
|---|---|
| **Large metro, central** | the entire population of the largest principal city of the MSA* |
| **Large metro, fringe** | Counties located in an MSA and with 1 million or more population |
| **Medium metro** | Population of 250,000 – 999,999 in an MSA |
| **Small metro** | Population of 50,000 – 249,999 in an MSA |
| **Nonmetropolitan** | Micropolitan (in an MSA) |
| **Nonmetropolitan** | Noncore (not in an MSA) |

* MSA = Metropolitan Statistical Area. The MSAs are defined by the OMB (Office of Management and Budget) as those places of having at least one urbanized area with a minimum population of 50,000. ** Census tracts = small geographic subdivisions of cities, counties, and adjacent areas. the average tract has a population of 4,000 residents and is designated to provide a degree of uniformity of population economic status and living conditions within each tract.

### The health of urban versus rural communities

Health differences have been reported between rural and urban communities, these differences reflect the different living conditions, infrastructure level, economic and overall access and quality to health care services. Overall, the differences between urban and rural population is probably not as dramatically different, but people living in rural areas are more likely to engage in negative behaviors that affect their health and quality of life. [26]

Differences in morbidity and mortality has been observed among urban and rural population in the United States. Some authors recently reported the trend one of the major disparity between urban and rural is in the issue of mortality, which persisted over time (the authors analyzed 47 years of data related to the problem, they found that the high level of mortality is tied to poverty more than any other associated factor.[27]. A general outline of the differences between urban and rural health is presented in the image below:[28].

26. Health Policy Institute. (n.d.). Rural and Urban Health. Georgetown University, McCourt School of Public Policy. From https://hpi.georgetown.edu/rural/
27. Cosby, A. G., McDoom-Echebiri, M. M., James, W., Khandekar, H., Brown, W., & Hanna, H. L. (2019). Growth and Persistence of Place-Based Mortality in the United States: The Rural Mortality Penalty. American journal of public health, 109(1), 155–162. From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6301407/
28. Difference Between. (June 5, 2021). Difference between Rural and Urban Areas. From https://finddifferencebetween.com/difference-between-rural-and-urban-areas/


Image from [Difference Between](#)

To illustrate in a more graphic manner, I found this infographic that is very useful.[29]

29. Gondi, S., Patel, K. (November/December 2016). Improving Rural Health. IEEE Pulse, a magazine of the IEEE Engineering in Medicine and Biology Society. From https://www.embs.org/pulse/articles/improving-rural-health/

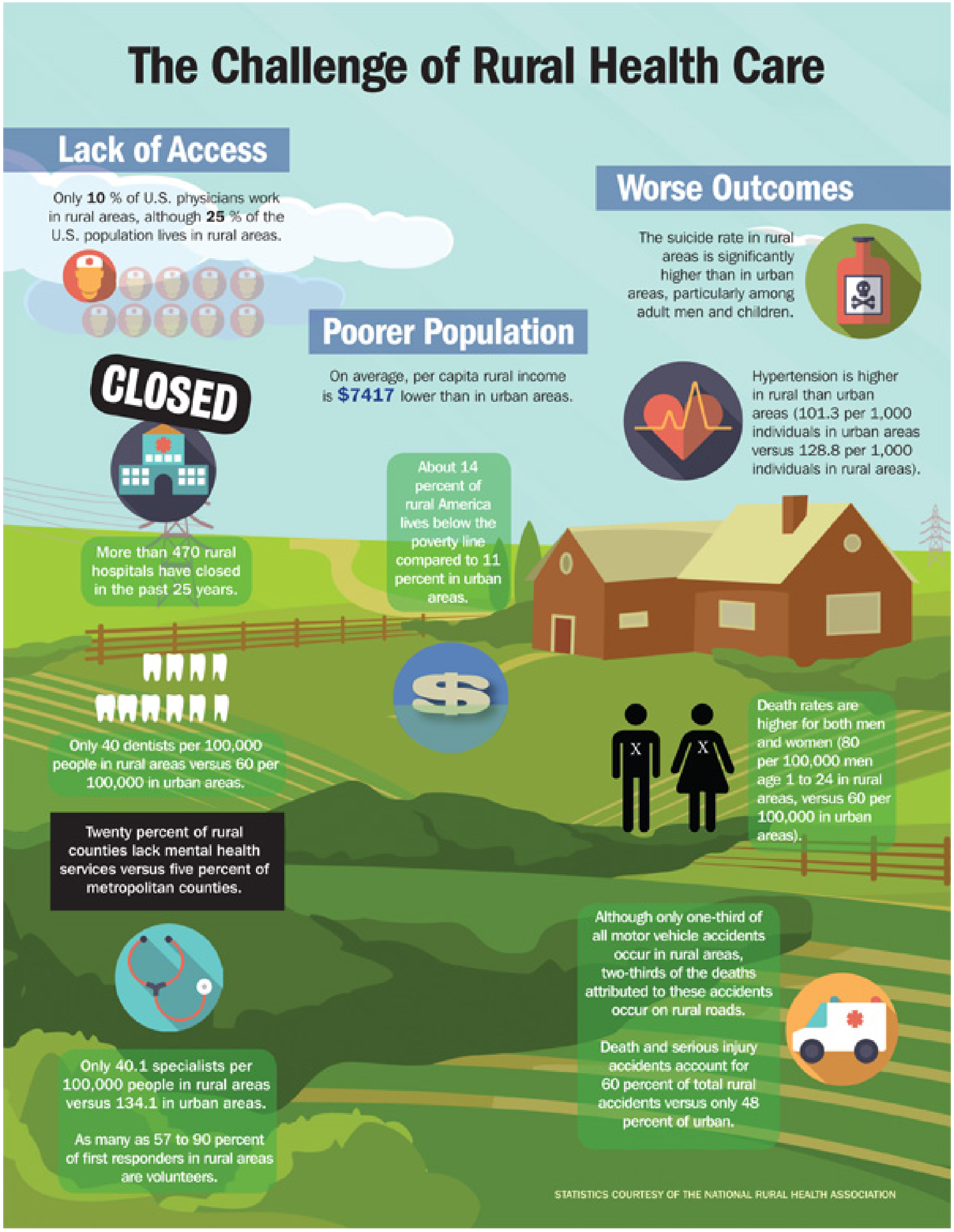



Image from [IEEE Pulse.](#)

**Variations of disease and other health outcomes within country**

The variation of disease and related health outcomes have been observed also within a same country geographical regions. For example, in the United States comparisons of disease rates have been established by region (Pacific, Central, Mountain, and Atlantic), and also by state. In the case of some chronic diseases such as cancer, it varies across states.[30] Similar situation has been reported on multiple sclerosis. Significant variations of multiple sclerosis rates have been found between north and south of the United States.[31] Also, similar variations can be found in the case of infectious diseases.[32]. A typical example of this is observed in the rates of Lyme disease in the U.S. in which the most affected areas are represented by the mid west states, and the northern east coast.[33]

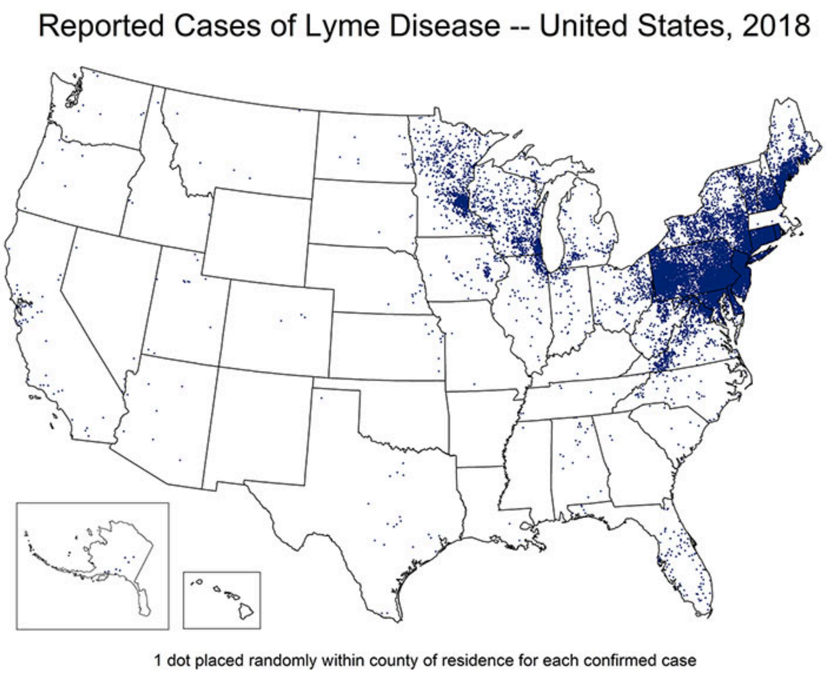



Image from [CDC](), [Lyme Disease Historical Data](), [Public Domain]().

30. Cook, L. (Oct. 15, 2015). You're Most Likely to Die From Cancer in 1 of These States Cancer death rates vary across the U.S. In US News. From https://www.usnews.com/news/blogs/data-mine/2015/10/15/youre-most-likely-to-die-from-cancer-in-1-of-these-states
31. Koskie, B., Sullivan, D. (August 21, 2020). Multiple Sclerosis: Facts, Statistics, and You. Healthline. From https://www.healthline.com/health/multiple-sclerosis/facts-statistics-infographic
32. Wilson M. E. (2010). Geography of infectious diseases. Infectious Diseases, 1055–1064. From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7152081/
33. CDC. (n.d.). Lyme Disease Maps: Historical Data. From https://www.cdc.gov/lyme/stats/maps.html

### Variation across countries

International variations across countries in the world show significant differences, and in some cases some similarities such as the presence of cardiovascular diseases that can be found around the world. In other cases, the differences are significant in terms of mortality, for example in the cases of the current COVID-19 pandemic, the differences across countries in the world varied, especially in those countries with poor health care system, high levels of poverty, and also a poor public health infrastructure among other factors. These mentioned conditions are expected to be found mainly in developing countries, however, the U.S. is not far away from those countries as a recent report from The Commonwealth Fund has found. This report shows the U.S. health care system is in general weak with among other indicators (see table below) has an administrative system that is inefficient. Here are the health care system performance of countries with similar levels of development and its rankings during the COVID-19 global pandemic.[34]

| | AUS | CAN | FRA | GER | NETH | NZ | NOR | SWE | SWIZ | UK | US |
|---|---|---|---|---|---|---|---|---|---|---|---|
| OVERALL RANKING | 3 | 10 | 8 | 5 | 2 | 6 | 1 | 7 | 9 | 4 | 1 |
| Access to Care | 8 | 9 | 7 | 3 | 1 | 5 | 2 | 6 | 10 | 4 | 11 |
| Care Process | 6 | 4 | 10 | 9 | 3 | 1 | 8 | 11 | 7 | 5 | 2 |
| Administrative Efficiency | 2 | 7 | 6 | 9 | 8 | 3 | 1 | 5 | 10 | 4 | 11 |
| Equity | 1 | 10 | 7 | 2 | 5 | 9 | 8 | 6 | 3 | 4 | 11 |
| Health Care Outcomes | 1 | 10 | 6 | 7 | 4 | 8 | 2 | 5 | 3 | 9 | 11 |

Image (table above) taken (with formatting modifications) "Health Care System Performance Rankings," from the Commonwealth Fund. Note: the score used in this study was developed using a ranking of 1-11, but there are more details in the **analysis of it.**

34. Schneider, EC., Shah, A., Doty MM., , Tikkanen, R., Fields, K., Williams II, RD. (August 4, 2021). Mirror, Mirror 2021: Reflecting Poorly Health Care in the U.S. Compared to Other High-Income Countries. The Commonwealth Fund, Fund Reports. From https://www.commonwealthfund.org/publications/fund-reports/2021/aug/mirror-mirror-2021-reflecting-poorly

Without going into details about each one of the criteria used by the authors of this publication, it can be seen that the U.S. is an outlier, it performed badly, its overall performance falls well-below the average for the rest of the developed countries listed in the report. The U.S. is last on all domains of performance except **care process**, on which it ranks second (or #2).[35]

### Other factors associated with variation in disease

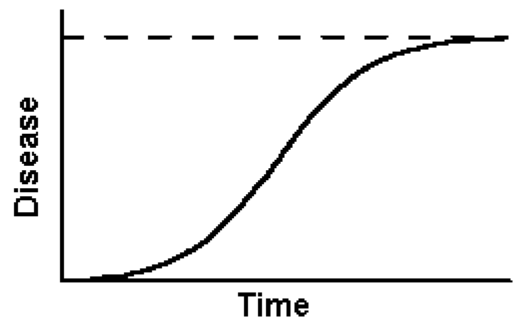
As the examples above have shown, the differences in the detection of disease is mainly link to location and geography. Other factors that may help to explain variations by place maybe the tendency of some ethnic groups to live in some specific areas of the United States, and also around the world. For example, Mormons tend to have a more frugal lifestyle than people living in Vegas, which accounts for variation of disease. Another factors could be social economic conditions and poverty levels found in some states in the country, for example, the high number of homeless persons in San Francisco, California; and New York city. Also, the levels of poverty in some states can explain the high levels of mental health problems found in very impoverished populations.

### Characteristics of Time

Following the who, where and when that we used in previous section of the chapter, time is the when, when did it happened? Adding the time dimension to the report of health phenomena (or, problems) makes a lot of sense since there are events in which the time is enough to bring the event to mind, for example, COVID-19 pandemic time, immediately the year 2020-21 comes to mind.

35. Comments from the health care rankings table found at https://infogram.com/mirror-mirror-2021-exhibit-1-1hdw2jp0qzlqj2l

So, if we know who (person) is affected, where (location) did this happened, and when, which is time; then, a picture of the health event is completed.[36]. In terms of a graphical representation of the concept of time, statistically speaking, the variable time belongs to the 'X' coordinate and the 'disease' to the 'Y,' as the image above represent.




A graphic representation of the relationship among time and disease. Image from apsnet.

**Other time concepts** In addition, it is customary (in most epidemiology books) to include in this section about time, the main characteristics of time, which includes the concepts of **cyclic variations, point epidemics, secular trends, and clustering.** All of these categories will be discussed in the content that follows.

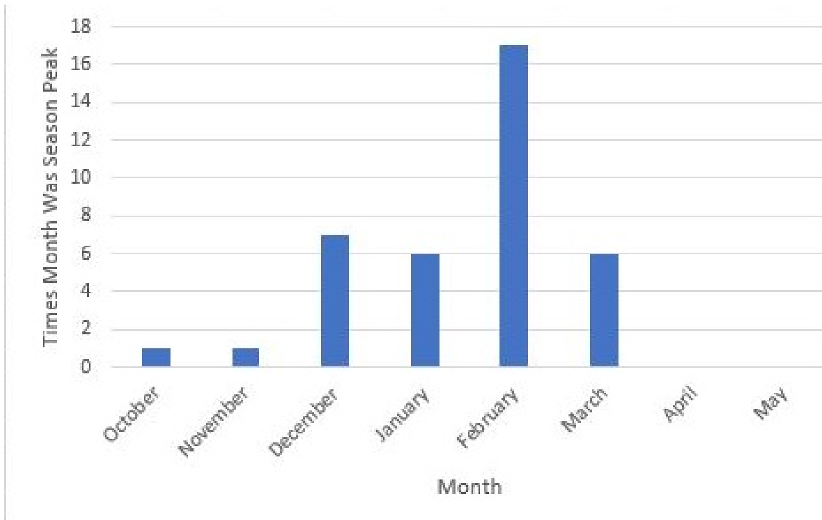
**Cyclic Fluctuations (or, cyclic variations)**

What are cyclic fluctuations? Taking the term directly from the Concise Encyclopedia of Statistics, "Cyclical fluctuations is a term used to describe oscillations that occur over long periods about the *secular trend* line or curve of a *time series*."[37] In other words, increases and decreases in the frequency of diseases and health condition over a period of years or within each year. Although it is not infrequent that these fluctuations (or, oscillations) reflect seasonal trends, they are not the same. So, how to identify cyclic fluctuations versus seasonal trends? Let's start by looking at the seasonal factors, these can be identified by winter and summer; the day of the week, the month or, the quarter of the year. So, for quantification purposes, *seasonal* health events are always reported *in a fixed and known period*. And, the periods are used to defined the time series which are the same as seasonal series. On the other hand, cyclic fluctuations appears in for

36. No author. (n.d.). Characteristics of Person, Place, and Time In PH717 Module 1B - Descriptive Tools, Descriptive Epidemiology & Descriptive Statistics. From https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module1B-DescriptiveStudies_and_Statistics/PH717-Module1B-DescriptiveStudies_and_Statistics3.html
37. No author. (2008) Cyclical Fluctuation. In: The Concise Encyclopedia of Statistics. Springer, New York, NY. From https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-32833-1_94#howtocite

example short period such as 2 years, and these fluctuations (increases and decreases, or, rise and falls) are of *not a fixed period.*[38]

**A common example of a seasonal disease is flu** which usually appears in the months of winter and early spring. most of the time flu activity peaks between December and February, but activity can last as late as May.[39] See image below:




United States flu season. From [CDC](). Image licensed [CC]() [SA]().

On the other hand, cyclic fluctuations frequently involved a disease that it does not appear (frequency) in specific times but appear and reappear in certain periods of time, for example, Pertussis (whooping cough), which peaks in disease every 3 to 5 years. [40]. See graph below:

38. Hyndman, RJ. (14 December 2011). Cyclic and seasonal time series. Definitions. Hyndsight blog. From https://robjhyndman.com/hyndsight/cyclicts/
39. Centers for Disease Control and Prevention. (n.d.). Flu Season. From https://www.cdc.gov/flu/about/season/flu-season.htm
40. Centers for Disease Control and Prevention. (n.d.). Pertussis (Whooping Cough) Frequently Asked Questions. From https://www.cdc.gov/pertussis/about/faqs.html

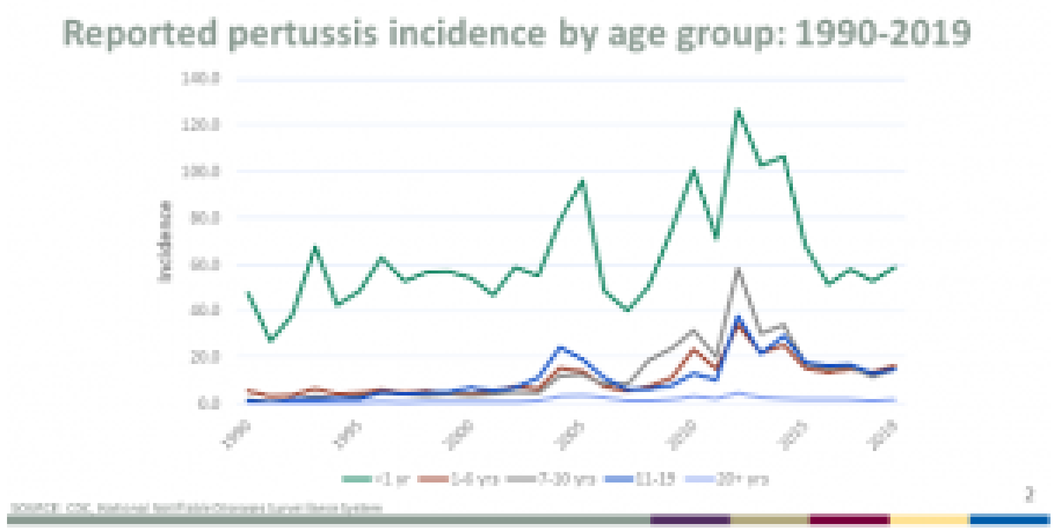
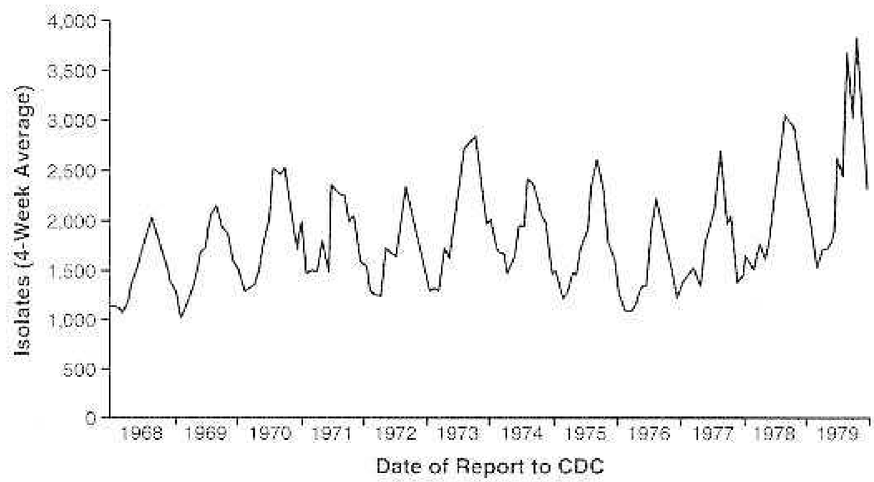



Image: Reported pertussis incidence by age group: 1990-2018. From [CDC.](#) Image licensed [CC](#) [SA](#). The data for this graph can be found at [CDC](#).

### Secular trends

The world secular comes from the Latin, “Saeculum”, that in general terms mean, someone who is not from the clergy, but in epidemiology, secular is used to refer to long period of times (usually years) in the occurring of some diseases. This trend is influenced by the degree of immunity in the population and possibly nonspecific factors such as poverty levels, or, lack of access to preventive health services.[41] An example is salmonellosis, see graph below:

41. Brachman PS. (1996). Epidemiology. In: Baron S, editor. Medical Microbiology, Chapter 9. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston. From: https://www.ncbi.nlm.nih.gov/books/NBK7993/




Reported human *Salmonella* isolations, by 4-week average, in the United States from 1968 to 1980. Source of the Image license not stated.

### Clustering

Cases about a disease can cluster in a group of individuals, closely grouped in time and place.[42] This can happen with diseases such as cancer. This cluster may have been linked to environmental exposures.[43] An image showing for example cancer clusters by state is shown below:

42. MedicineNet. (3/29/2021). Medical Definition of Cluster. From https://www.medicinenet.com/cluster/definition.htm
43. Goodman M, Naiman JS, Goodman D, LaKind JS. (2012). Cancer clusters in the USA: what do the last twenty years of state and federal investigations tell us? Crit Rev Toxicol, 42(6):474-90. From https://pubmed.ncbi.nlm.nih.gov/22519802/

Geographical clusters of U.S. counties with significant high or low... | Download Scientific Diagram

Geographical clusters of US counties with significant high or low breast-cancer. Source, Licensed CC BY 2.0

Note: There are more content that can be covered under the topic discussed in this chapter, but for now, the book will cover only what it has been written until here.

**Summary**

This chapter has introduced the concepts of person, place and time commonly used in epidemiology. The study of these three categories is not as simple as it appears, there are several sub-categories that exist for each one of the main topics, and it is the intention of the author that the images, and examples helped to make sense of the information.

# 5. Study Designs Commonly used in Epidemiology

## Study Designs Commonly used in Epidemiology

*Learning Objectives*

By the end of this chapter, the learner will be able to

- Describe the most common research study designs used in epidemiology
- Differentiate between Non-experimental Observational studies, and Experimental/ Interventional epidemiological studies
- Differentiate among individual and population based studies, and also between observational, descriptive and analytic studies.
- Understand the use of randomization in experimental studies such as clinical trials, and other types of experimental field trials.

### Introduction to the chapter

This chapter will present the most commonly used epidemiological study designs,listing main characteristics and then, focusing on their benefits, strengths, weaknesses, and uses in public health.

Most epidemiologists are trained to do their investigation based on a series of designs called, **Study Designs.** The study Designs commonly used in epidemiology are based on several premises, but a series of questions can help the investigator to decide what design best fits its needs, some of these questions are, what types of study designs are there? How and when we use specific type of study designs? Which study design is the most appropriate to use in certain investigations? The list of questions could continue but it is important to generate these questions in order to arrive to a decision of what fits better the investigators needs. Also, investigators need to be familiar with these study designs so they can use them when needed.

*Main Question*

**What are the most common study designs used in epidemiology?** The answer to this question is this chapter, so, you are in the right place, continue reading...

**What are study designs in epidemiology?** Study designs refer to the different approaches mainly used to conduct research for investigative purposes. They are called, 'designs' because they represent a specific manner of conducting the research process, which is mainly based on the scientific method. Study designs are more of a framework to guide the researcher in the process. [1] And, although basically all research process starts with a research question, there is need to follow a process that will convert this research question into a hypothesis, and then, to a real life situation, or, scenario that need to be framed in order to arrive to valid conclusions. [2] In other words, study designs assists the researcher providing a type of road map that will help to not get lost. It is easy to get lost, especially when complex health phenomena is studied/researched, but the study design is expected to assist in providing direction. [3] In sum, study designs are road maps, or, frameworks that assist in the research/investigative process.

It is also probably useful to mention that some of the study designs mentioned here, are not only used in the sciences of epidemiology, they are used by other areas of study, especially those areas that belong to the social sciences, including public health, but also mathematics, statistics, and of course, medical sciences. So, this study designs are not unique to the field of epidemiology, but they are highly used in public health, and medical research.

### Descriptive versus Analytic

In a broadly manner, epidemiologic study designs can be divided into two broad categories: 1) Descriptive and 2) Analytic.

**Descriptive studies** as the word implies, 'describe' situations, problems, and other

1. No Author. (n.d.). Introduction to study designs - geographical studies, available at: https://www.healthknowledge.org.uk/e-learning/epidemiology/practitioners/introduction-study-design-gs
2. Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: advancing the methodology. Implementation science : IS, 5, 69. https://doi.org/10.1186/1748-5908-5-69
3. P Sai Kumar, Imperial College London.(n.d.). Epidemiology for Practitioners. Resource text adapted from material written by Maria Kirwan, available at www.healthknowledge.org.uk

health phenomena (diseases, disorders, health behavior, healthy lifestyles, etc.). These type of study designs are mainly used to generate hypotheses, especially in those cases in which the health issue in study is unknown, or, there is not much information on it, or, simply, because the topic is a 'new' problem found in sciences, in this case, social sciences, including epidemiology. On the other hand, **Analytic studies,** 'analyze' the health phenomena, situation, or, problem.

**More elements used to distinguish between descriptive and analytic research studies**

Since epidemiology is by nature quantitative, the division between descriptive and analytic studies can be also clearly recognized by the type of quantitative methods, in this case, the methodology, data collection, and statistical analyses that are used in the research process. For example, in descriptive study designs, the most common data analysis is the use of 'descriptive statistics' such as numbers, percentages, sums (total number of cases), mean, mode, standard deviation, etc. Since, the descriptive study is looking mainly for a 'description' of the problem, the use of descriptive analyses suffice for the type of research that mainly intents to generate hypotheses, or, to add more information for future studies.

**Overview of Study Designs** – the following is a list of study designs:

***Descriptive Studies***

In this category, the following are the most commonly used/ listed:

a) Case Study or, Reports
b) Case Series
c) Cross-sectional studies
d) Ecologic studies

***Analytics Studies***, since they are more complex, they are subdivided into two additional categories: *Observational and Experimental.*

> ***Overview***
>
> The content in this section of the chapter intends to provide a general overview (or, a panoramic) of the types of study designs commonly used in epidemiology.

**Observational studies** are mainly represented by the following:

a) Cross-sectional studies – as you probably had noted, this category is shared also

with the ‘descriptive’ category, which means that a cross-sectional study can be descriptive, observational, or, analytic.
b) Case- control studies
c) Cohort studies

**Experimental studies**
In this category, it is customary to include the following:
a) Clinical Trials
b) Community Trials
c) Other forms of experimental research.

**The information listed above can be summarized in the following Figure**

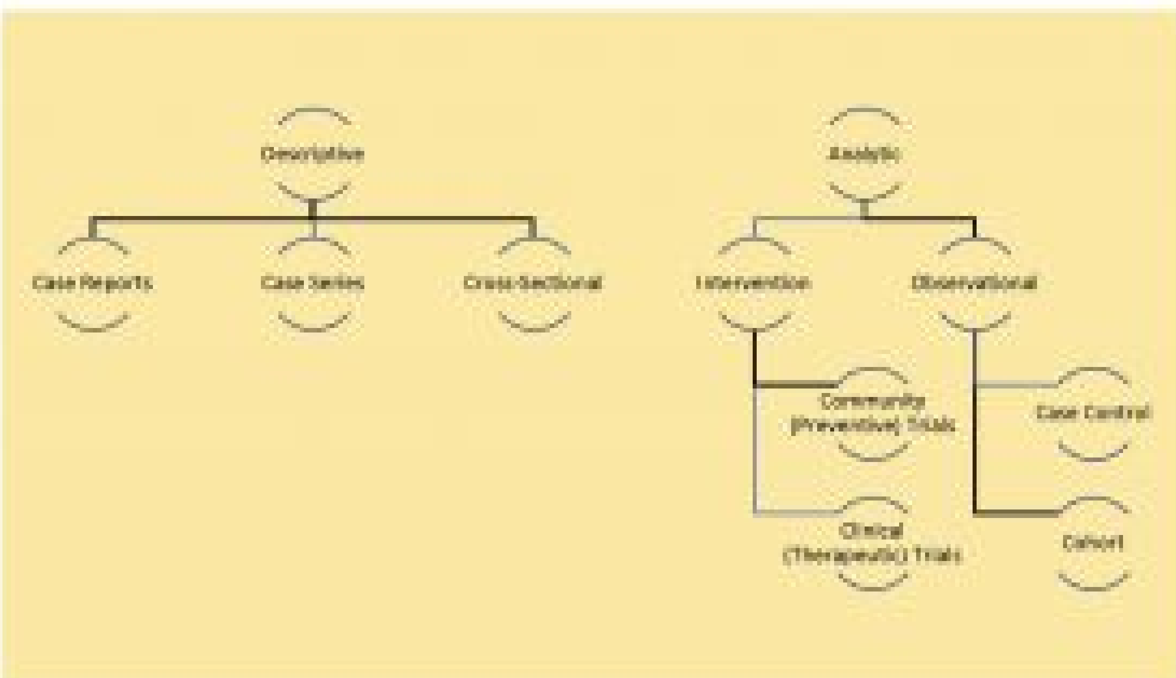



Figure 1. Types of Epidemiological Studies. Image prepared by Giovanni Antunez, Licensed CC BY 4.0

The information has been summarized somehow in publications, such as the ‘Study designs commonly used in epidemiology, from community medicine, study designs, howmed net.[4]

Now, the question is, **how do we know what study design to use?** As it is seen (listed) above, there is a repertoire of study designs available to the investigator. The use of any of these designs depends on the mainly purpose/reason for the study, and also, the resources, which are mainly financial, and expertise of the researcher team.

*Question*

**how do we know what study design to use?** The answer is here in this section of the text.

**How do we choose certain study design from the list of options?**

Many times, the answer to this question depends on the *purposes/reasons* for the research, or, investigation of a certain health phenomena. For example, if we want to conduct a study about a topic/area that is unknown, or, not well-known, and the research question is still a work in process, or, it is clear, but there is need to elaborate a hypothesis/or, hypotheses; then, the most appropriate study design in this case is the descriptive, why? It is because descriptive studies are commonly used for hypotheses generating. Does it mean that there is no other way to conduct the study? Not necessarily, because they may be another way to do the same, but over and over, especially in social sciences, public health, and specifically in epidemiology, descriptive study designs have been used for those purposes, in this case, hypotheses generating.

Another example that refers to the purposes/reasons for the research but that also refers to the financial aspects of medical, and public health research is the cohort study. In this case, if the main reason/purpose of the study is to find a causal relationship, the cohort study is the answer. This type of study design will help overtime to elucidate the associated risk factors, and social health determinants that are related to the health problem that is researched. What is the only reason that stop the use of the cohort study? The main obstacle is financial, cohort studies can be highly costly, since, study subjects are basically follow over a long period of time until the health outcome is developed, as it is the case of for example the cardiovascular disease cohort studies conducted in the United States. More

4. community medicine, study designs, howmed net.: HYPERLINK "http://howmed.net/community-medicine/study-designs/" o "Figure: Study designs" t "_blank" http://howmed.net/community-medicine/study-designs/

information about cohort studies will be provided/expanded later in the context of this chapter, and the overall book content itself.[5]

## Descriptive Study Designs

### The Case Studies/Case Series

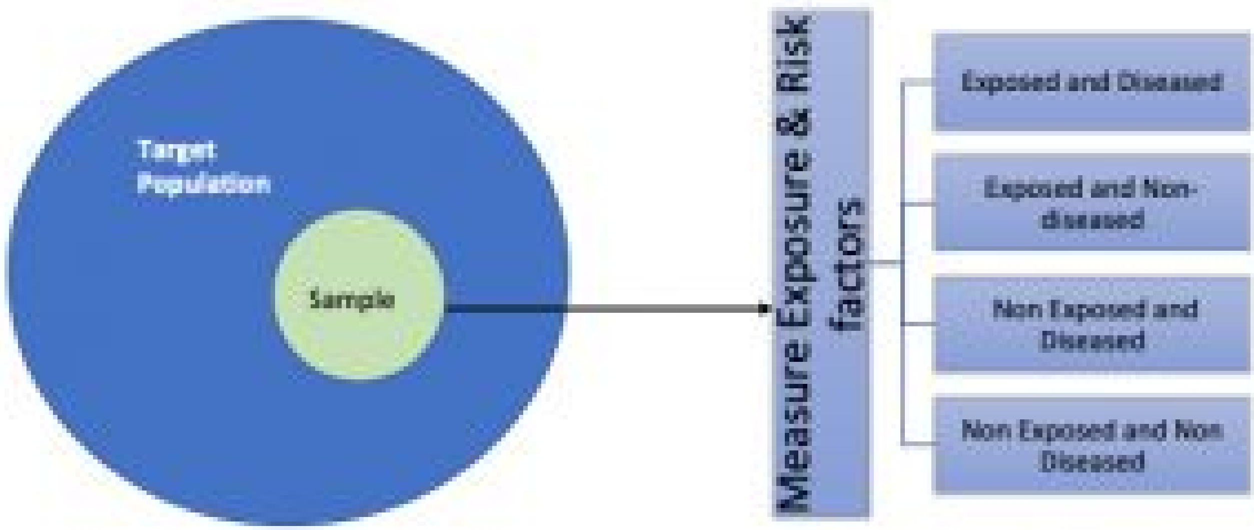
This study design is commonly used when there is no much information about the case, as it is the example of a recently reported disease, or, a disease that is very rare, so, the investigator wants to share the information with the scientific community, but since there is only one, or, three cases, it is much better to choose this study design, which sometimes become a case series, if a continuation of cases are reported in a limited fashion. The main limitation is that the cases are not necessarily representative of the general population, but the benefits are that the reported case brings an opportunity for future studies on the subject.[6],[7]

In most epidemiological textbooks, the case studies are commonly used by the medical community to report a case, or, a series of cases that usually represent patients suffering from certain diseases. The medical model usually reports cases that are diseased. But the model hardly applied to other health phenomena, and since it is usually oriented to report cases in a limited fashion, there is small use of this design in epidemiology, which focuses on populations at large.

### Ecologic Studies

For this study design, the unit of analysis is the group, not the individual. In this case, correlations are obtained between exposures rates and disease rates among different groups, or, populations. Because of the word, ‘ecologic’, ecological studies tend to be confused with ‘environmental’ studies, and this is possible especially if an environmental issue is studied using this design, but in general, ecologic studies refer to the group(s) investigated.

5. Georgia State University Library. (n.d.). Research Guides, Literature Reviews and Types of Clinical Study Designs. From https://research.library.gsu.edu/c.php?g=115595&p=755213 Last Updated: Mar 11, 2022 2:42 PM
6. Deakin University Library. (n.d.). Quantitative Study Designs: Case Study / Case Report / Case Series. From https://deakin.libguides.com/quantitative-study-designs/casestudy
7. Kooistra, B., Dijkman, B., Einhorn, TA., Bhandari, M. (May 01, 2009). How to Design a Good Case Series, The Journal of Bone & Joint Surgery: 91(3), 21-26. From https://journals.lww.com/jbjsjournal/fulltext/2009/05003/How_to_Design_a_Good_Case_Series.5.aspx doi: 10.2106/JBJS.H.01573

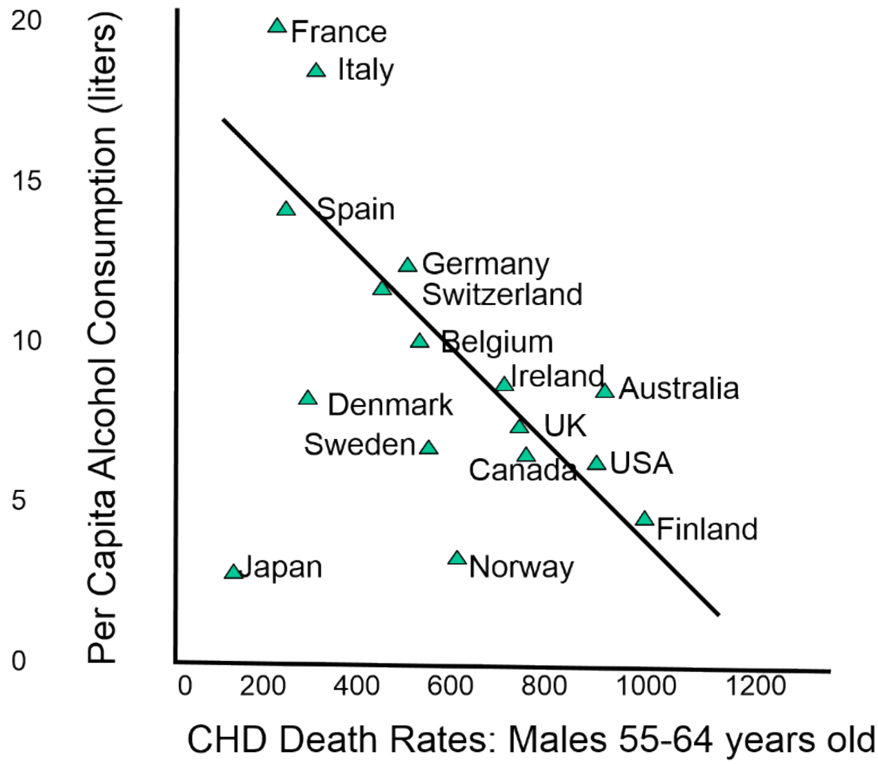



Ecologic Study Design. Image created by Giovanni Antunez, CC BY 4.0

The ecological study provides a setting in which observations made at the group level may not represent the exposure-disease relationship at the individual level, this is called, the *ecologic fallacy*, which occurs when incorrect inferences about the individual are made from the group level data. [8].

Essentially, this means that inferences from 'the results of ecological studies can only be applied to the group but not to the individual'. You may said, why? And the answer is because the intention of the ecologic studies is to capture how the health events are affecting the group, the community and not the individual. An example of this, is this study done in alcohol consumption and coronary heart disease (CHD), see the image below that presents the respective information:

8. Freedman, DA. (199). Ecological Inference and the Ecological Fallacy. Department of Statistics, University of California Berkeley, available at https://web.stanford.edu/class/ed260/freedman549.pdf




'Death rates of coronary heart disease (CHD) in different countries' , image from PH717 Module 1B

In the study of the image presented above, the data across countries (the 'population' or, 'group' mentioned in the definition of ecologic studies) showed that moderate use of alcohol was not beneficial to the heart, on the contrary, it increases the risk for CHD, why? It is the typical case in which what is true at the group or, population level, it is not true to the individual level. Studies done on individuals or, specific groups had shown that moderate alcohol consumption is beneficial to prevent CHD.[9]

As it has been presented in the example above, the ***ecologic fallacy*** which is part of the nature of an ecologic study, it is also considered a disadvantage of this type of

9. LaMorte, W.W. (2020). Death rates of coronary heart disease (CHD) in different countries In PH717 Module 1B, Descriptive Tools, Descriptive Epidemiology & Descriptive Statistics. Boston University School of Public Health. From https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module1B-DescriptiveStudies_and_Statistics/PH717-Module1B-DescriptiveStudies_and_Statistics6.html

study design. Another disadvantage is that ecologic studies could make imprecise measurement of exposure and disease. [10].

On the other hand, the following are advantages of ecologic studies, they are quick, simple and less costly than other studies, and their completion is faster compared to other designs used in epidemiology and related sciences. One more advantage is that they can be used (and very useful in this sense) for generating hypotheses, especially when a disease is of unknown etiology.

**Common uses of ecologic studies**

Specific applications of the ecologic study design has been classified by some authors as the following: geographical comparisons, time trends, migrants, occupation and social class. More details are included below:

*Geographical comparisons,* which for example can be used to find prevalence of risk factors by comparing incidence or, mortality in two, or, more geographic areas. *Time trends,* which essentially means to study the fluctuations on the incidence of chronic diseases which tend to change over time. *Migrants*, the study of migrant groups can be used to identify those factors that are predominantly genetic from those who are environmental, in this case, first or, second generation of an ethnic group maybe affected differently depending on their degree acculturation. *Occupation and social class,* in this case, *it refers on how some morbidity and mortality area associated with certain occupations, and also with the socioeconomic status of the groups working on those type of jobs.*[11]

## Cross-Sectional Studies

This study is a commonly used design. As a way to understand the most basic principle of a cross sectional study, is to think on the total study population as a pie, in which each percentage represents a section of the pie, then, for study purposes only a piece (or, section) of the pie is investigated. Since in the majority of cases, the characteristics of a population are very similar, choosing to study one portion

10. Freedman, DA. (199). Ecological Inference and the Ecological Fallacy. Department of Statistics, University of California Berkeley, available at https://web.stanford.edu/class/ed260/freedman549.pdf
11. British Medical Journal (BMJ). (n.d.). Chapter 6. Ecological studies (Geographical Comparisons, Time Trends, Migrants, Occupation and Social Class) in Epidemiology for the uninitiated. From https://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/6-ecological-studies#chapters

(section) of the pie (the population) will be representative of the total population. Assuming that we are talking about a study population, not the general population. This analogy takes us to the term, cross (cutting) section (a piece of the pie) study.

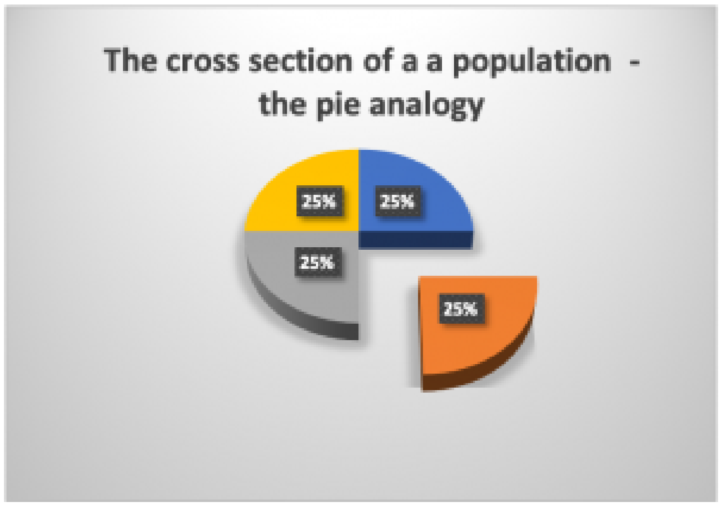
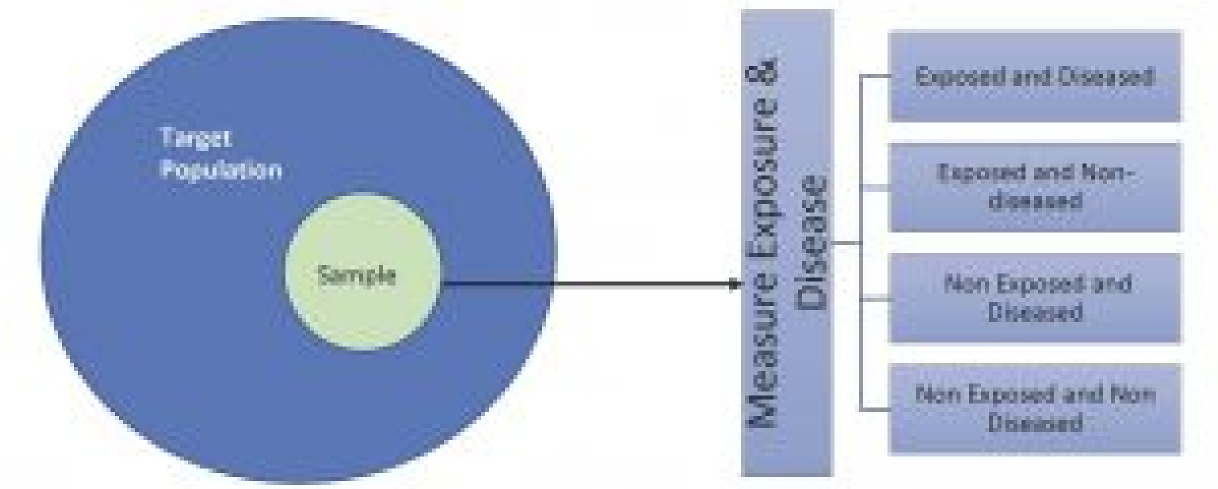



**Figure 2.** The Pie analogy to understand the concept of Cross-Sectional study. Figure prepared by Giovanni Antunez. [Licensed CC BY 4.0]



### More about the cross sectional study design

Keeping the above analogy in mind, let's look at the cross-sectional study in another way, first we draw a single sample from the target population and assess current exposure and disease status on everyone, see image below:




Cross Sectional Study Design. Image created by Giovanni Antunez, CC BY 4.0

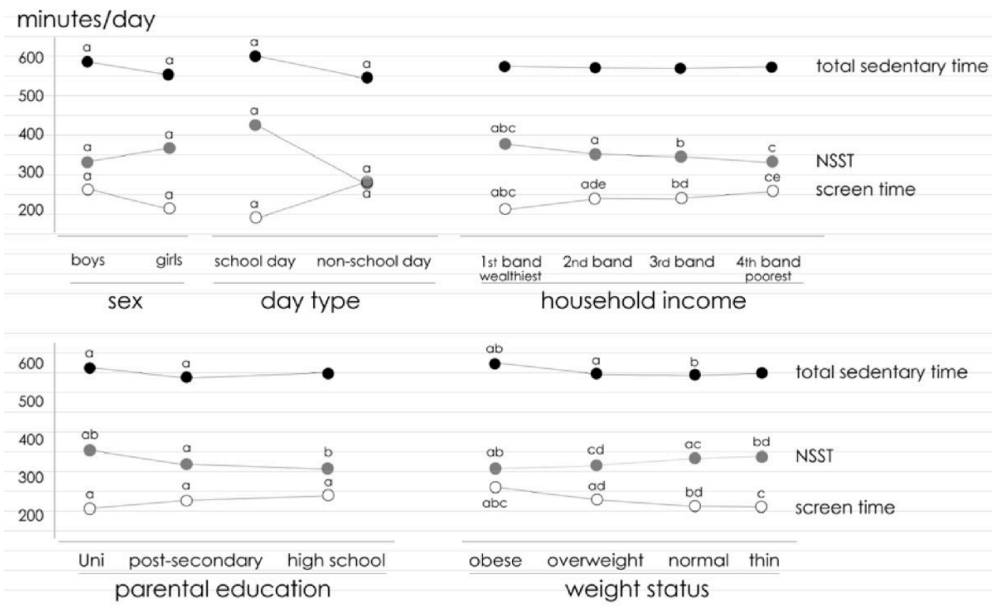
Since cross-sectional studies collect data at a point in time, they are commonly used to calculate prevalence for public health reports, and also for the designing and location of health services in a community. This study design is used frequently, especially when there is not much money to afford another type of study. So, cross sectional studies are very popular not only because they are less costly, but also they are fast to complete. Another great advantage of cross sectional studies is that they are used to generate hypotheses, or, specific research questions about exposure and disease, however, this type of study does not address the issue of **temporality**, due to being one shot only (one point in time), it does not provide information to know what was first between cause and effect.[12]

### Example

An example of this type of study design is, an Australian cross-sectional study on the effects of screen and non-screen sedentary time in adolescents, and how these types of behaviors affect their weight and overall well-being. The study found that although screen sedentary time (SST) is a contributing factor in the amount of

12. Setia M. S. (2016). Methodology Series Module 3: Cross-sectional Studies. Indian journal of dermatology, 61(3), 261–264. https://doi.org/10.4103/0019-5154.182410

fatness observed in school-age children and adolescents, there are two other factors that need to be studied if a significant change is expected to be observed, and these are active lesson breaks in the classroom, or active transport to school. These two last factors are part of what the study called, the NSST or, Non-screen sedentary time.[13] As part of the study design (cross-sectional), some of the study variables are presented in the figure below (taken directly from the published article), see below:



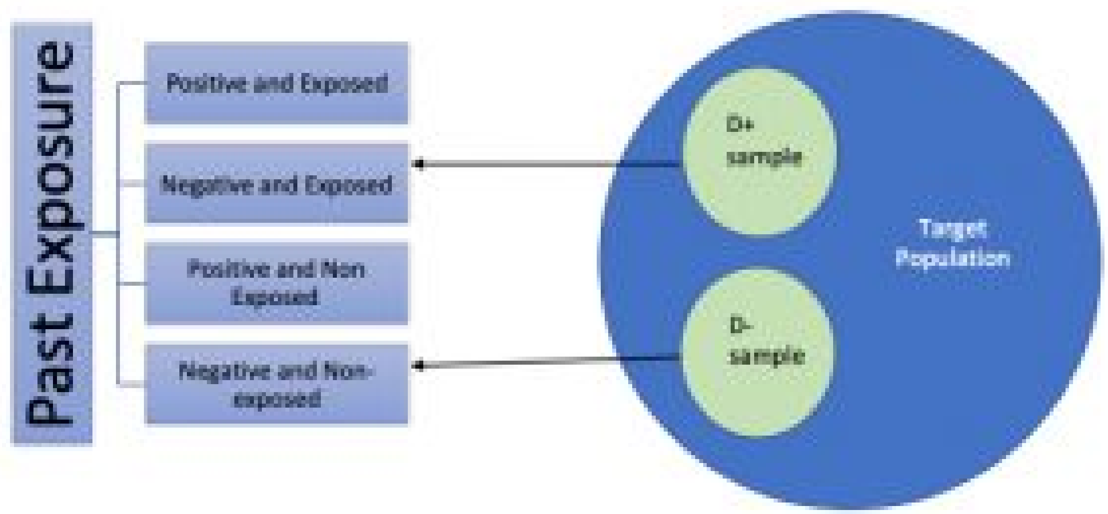

'The relationship between non-screen, screen and total sedentary time' [selected variables] – example of a cross-sectional study, image from the article cited/mentioned above.

Finally, one of the major benefits of cross sectional studies is they are quick (compared to other type of designs such as the cohort study) to complete, making them very efficient in terms of time and cost.

13. Olds, T.S., Maher, C.A., Ridley, K. et al. (2010). Descriptive epidemiology of screen and non-screen sedentary time in adolescents: a cross sectional study. Int J Behav Nutr Phys Act 7(92). From https://doi.org/10.1186/1479-5868-7-92

**Case-control studies**

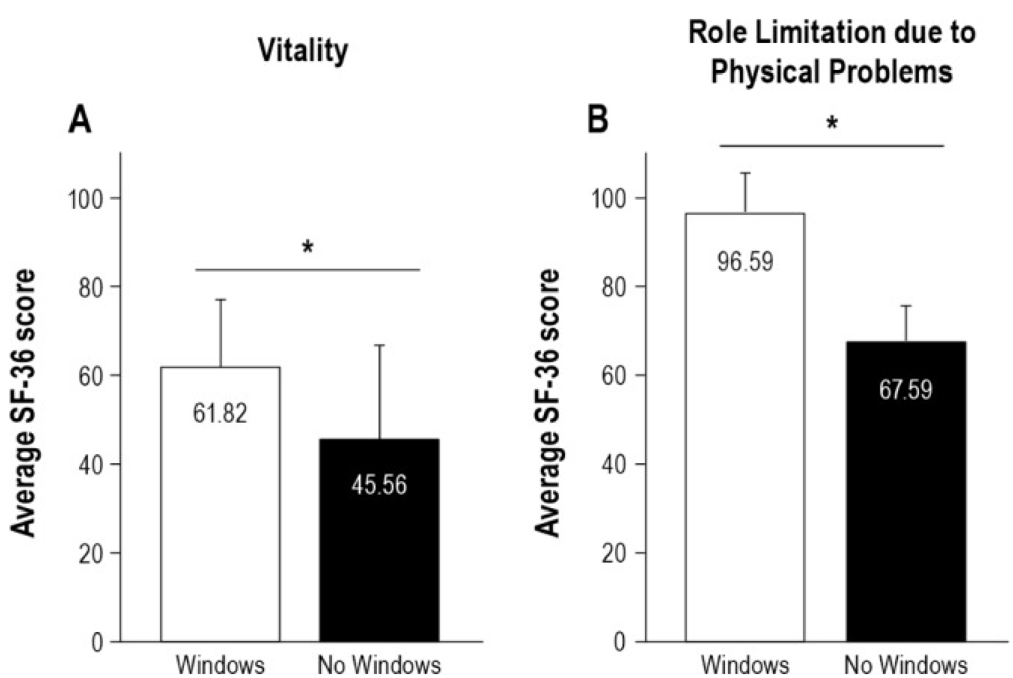
This type of study design is a relatively commonly used study, and the main reason is because analysis the exposure and disease event in a retrospective manner (the data has been already collected, and it can be found from medical records, disease registers, and databases of some specific diseases/health problems. Below is a graphical representation on how the cases and control are selected:




Case Control Study Design. Image created by Giovanni Antunez. CC BY 4.0

The study process requires, the identification of the individuals affected by the disease in study, which are called, 'the cases,' and their counterparts who are similar in characteristics (mainly demographics), which in this case are called, 'control.' So, a comparison between cases and controls is used to assess the risk associated with the development of a disease or, health problem. Since the data on the exposed (the cases), and the non-exposed (the controls) had happened in the past, the case-control study design is useful to assess prevalence but not incidence. Another characteristic for this type of study design is, that similar to the cross-sectional studies, the data collection is a single point in time, with the exception that in the case of case-control studies, the main characteristic is that the selection of the study subjects (participating) is based on the presence or, absence of the outcome.[14]

**An example** of a case control study is the work on a group of investigators who compared the Impact of windows and daylight exposure on overall health and sleep quality of office workers. The results showed that workers in windowless environments tend to experience limitations in their role in terms of physical problems and vitality and some sleep disturbances. When the two groups were compared, workers with windows had more light exposure, more physical activity, and longer sleep duration.[15] A selected image included in the article shows graphically how two (more variables were studied) of the characteristics between both groups show clearly the difference that makes to have windows or, not in the workplace.




'Windows and no windows in the workplace', Image from the article cited/mentioned above.

14. Tenny, S., Kerndt, C. C., & Hoffman, M. R. (2022). Case Control Studies. In StatPearls. StatPearls Publishing. From https://pubmed.ncbi.nlm.nih.gov/28846237/
15. C. H., & Zee, P. C. (2014). Impact of windows and daylight exposure on overall health and sleep quality of office workers: a case-control pilot study. Journal of clinical sleep medicine, 10(6), 603–611. From https://doi.org/10.5664/jcsm.3780

### Common uses of the case-control study design

Because of all of the mentioned characteristics, the case control-study design is used to find the prevalence in the community of certain diseases, and from their results, a subsequent study is designed. Also, the case-control study design has been used especially for the study of rare diseases.

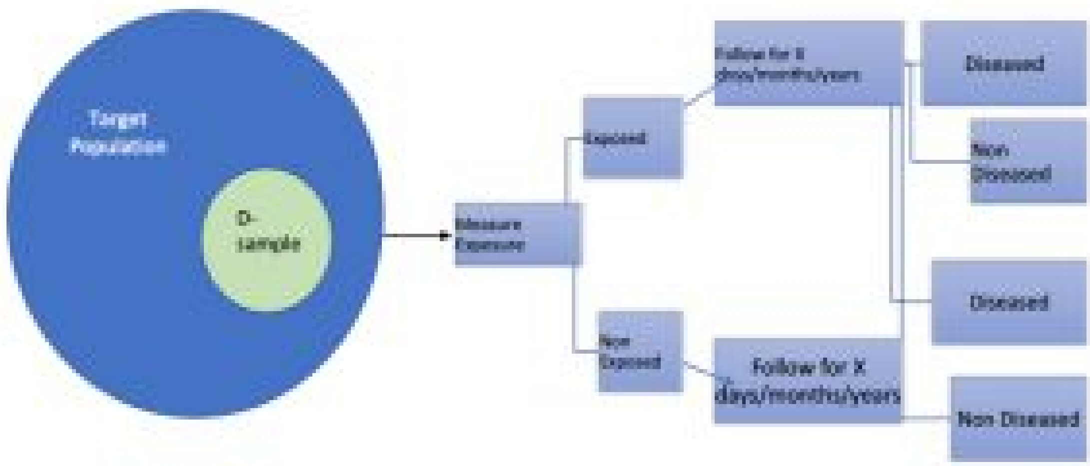
### The selection of the cases and the controls in the case-control study design

This is an important step in the use of the case-control study design. The cases need to be selected based on a set of criteria, which defines the characteristics and manifestation of the disease, including laboratory and other medical tests such as imaging, including x-rays. Then, these criteria is used to identified the cases. And, what about the controls, how are those individuals found? The controls can be for example patients from the same clinic/hospitals as the cases, or, the same population, for example, college campus, or, factory workers. The major characteristic of the controls is the similarity in terms of the selection criteria used for the cases, so, a comparison can be established. When there is time to assess the risk, both groups cases and controls are included in the statistical calculations based on the exposed, and not exposed criteria and the development of the disease under study. It is important to note that the word, 'exposed' here is a matter of semantics, because the exposed do not necessarily existed, they are cases (they have the disease already).[16]

### Cohort Study

This type of study design is considered the prototype (the model) of an almost 'perfect' design to investigate causality. In the cohort study, the main measure of disease frequency is, incidence. The following diagram presents how the study population is selected, and the major steps in the implementation of this type of study design:

16. Setia M. S. (2016). Methodology Series Module 2: Case-control Studies. Indian journal of dermatology, 61(2), 146–151. https://doi.org/10.4103/0019-5154.177773




Cohort Study Design. Image created by Giovanni Antunez. CC BY 4.0

In the field of epidemiology is also accepted that cohort study address the issues of *temporality* (most studies do not) making possible to avoid logical errors. Cohort studies are usually conducted for longer periods of time compared to other designs; and the reason for it is that cohorts start with individuals who are free of the disease under study, and are followed up over the years to observe the development of this disease (or, group of diseases such as cardiovascular diseases). Due to the fact that most of the time, the data is collected in the future (from one specific point in time – the now, and the upcoming time of observation), the word retrospective is commonly used to reflect this concept, which makes most cohorts, examples of perspective studies.[17]

In some cases, a cohort study can be designed by using data that has already been collected (which resembles the case-control design), and since this is data collected in the past, then, the cohort is called, a *retrospective cohort*. And, when, retrospective data, present time data, and prospective data collection (which is essentially the classic cohort study) are included, the name of the cohort is,

17. Song, J. W., & Chung, K. C. (2010). Observational studies: cohort and case-control studies. Plastic and reconstructive surgery, 126(6), 2234–2242. https://doi.org/10.1097/PRS.0b013e3181f44abc

*ambispective*. But in reality most people when they hear the word, cohort, they are referring to prospective data collection studies (or, prospective cohorts).[18]

Cohort studies can be used to study more than one disease, or, multiple exposures, so, investigators can take some data (already collected in the cohort), and design a case-control study known as the *'nested case-control,'* it is nested because it comes from inside the cohort.[19], [20]

Because of all of the mentioned benefits and advantages of cohorts, especially the calculation of incidence about a disease, makes cohorts an ideal study design, but at the same time it limitation is that cohorts are highly expensive, and that is mainly due to the fact that they last for a long time, especially for those diseases that take a long time to develop, so, the investigators have to wait for a while before they see the first generation of cases.[21]

**Examples of Cohort Studies**

To provide a mental picture of the cohort study design, I am including here, what I called, 'Famous Cohorts in the United States.' Famous because they reflect the reality of the country in terms of race segregation, and other socio-demographic factors that have shaped the country.

### The Framingham Heart Study

The Framingham Heart Study takes its name from the town in which the study was conducted, Framingham. Framingham is located in Massachusetts, United States, and it within Middlesex County and the MetroWest subregion of the Greater Boston metropolitan area. [22] A picture of this city is shown below:

18. School of Public Health, Boston University. (n.d.). Cohort Studies. From https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_cohortstudies/ep713_cohortstudies_print.html
19. Ernster V. L. (1994). Nested case-control studies. Preventive medicine, 23(5), 587–590. https://doi.org/10.1006/pmed.1994.1093
20. Soyoung K. (2016). Case-Cohort Studies vs Nested Case-Control Studies. Datum, Newsletter of Biostatistics, Medical College of Wisconsin (MCW). 22(1). From https://www.mcw.edu/-/media/MCW/Departments/Biostatistics/vol22no1kim.pdf?la=en
21. School of Public Health, Boston University. (n.d.). Advantages & Disadvantages of Cohort Studies. From https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_cohortstudies/EP713_CohortStudies5.html
22. Framingham, Massachusetts, Wikipedia, From https://en.wikipedia.org/wiki/Framingham,_Massachusetts

The


‘Odd Fellows Building , Framingham MA’, image from Wikimedia Commons, licensed CC 3.0 Unported.

Framingham Heart Study is a classic example of a cohort study that assessed multiple exposures and multiple outcomes. This study, a collaboration between the US National Heart, Lung, and Blood Institute (a division of the National Institutes of Health) and Boston University, began in 1948 by enrolling just over 5,000 adults living in Framingham, Massachusetts. Investigators measured numerous exposures and outcomes, then repeated the measurements every few years. As the cohort aged, their spouses, children, children's spouses, and grandchildren have been enrolled.[23]

The Framingham study is responsible for much of our knowledge about heart disease, stroke, and related disorders, as well as of the intergenerational effects of some lifestyle habits.[24] More information and a list of additional publications (more than 3,500 studies have been published using Framingham data) can be found extensively and especially in the project's website.[25]

'Framingham Heart Study Physicians, 1948',
Image from Flickr.

**The Bogalusa Heart Study**

Although, the Framingham Heart Study is a model cohort that influenced the work in public health and medicine. There was one flaw with the study, the participants were all white or, Caucasian; which from the beginning introduced a confounding factor that is race, which is a health determinant that is also linked to income, socio-economic status, social class, and among others, access to health services. So, to study cardiovascular disease beyond the white population generated the need to conduct a study on another major ethnic population group in the U.S. population,

23. No Author. (n.d.). Framingham Heart Study, Participant cohorts. Available at https://framinghamheartstudy.org/participants/participant-cohorts/
24. Framingham Heart Study. (2019). Participant Cohorts. From https://framinghamheartstudy.org/participants/participant-cohorts/
25. Framingham Heart Study. (n.d.) The Framingham Heart Study, Available at http://www.framinghamheartstudy.org/

which is the black or African American community. Although not many studies on black, the Bogalusa Heart Study in Louisiana was born in 1972. Bogalusa is a small town in Louisiana (almost in the limit with Mississippi), it is mainly a biracial (black/white) rural community in which entire families have lived there for generations. The population in Bogalusa is mostly constant with few, or, no migration is ideal study population brought the attention of a famous Tulane University School of Public Health in New Orleans, Dr. Berenson, who lived through the entire duration of the study. [26], [27]. See a picture of the town, and also of Dr. Berenson:

"Bogalusa City Hall". Wikimedia Commons. Licensed CC BY-SA 4.0

'Dr. Gerald Berenson' the founder of the Bogalusa Heart Study, image from American Heart Association Journal.

He survived the study which was taken down after hurricane Katrina due to major damage by the lack of electricity in New Orleans during Katrina in which many of the study samples that were stored were damaged, and also, the lack of funding after the impacts of Hurricane Katrina devastation in New Orleans in 2005.

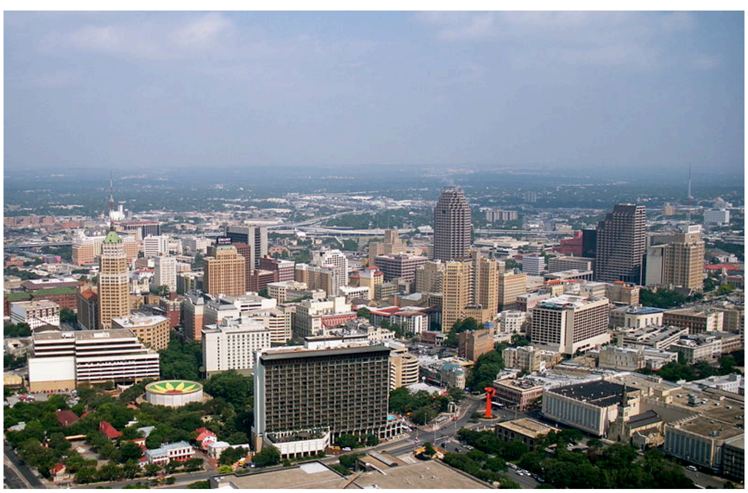
**The Bogalusa Heart Study** started as an epidemiological study of cardiovascular risk factors in children and adolescents; it eventually evolved into observations of young adults. This study *main milestones confirmed* the findings of the Framingham Heart Study, but also superseed in terms of adding new variables to the study of

26. Wikipedia. (n.d.). Gerald Berenson. From https://en.wikipedia.org/wiki/Gerald_Berenson
27. No Author. (n.d.). Bogalusa Heart Study. Available at: https://www.clersite.org/bogalusaheartstudy/

cardiovascular disease, especially with the findings of the presence of cardiovascular disease in children, which had not been studied before. The study reported an African American child who died of cardiovascular disease and had atherosclerotic deposits in his arteries at the age of eight years old. This finding moved the American Heart Association to recommend that children stop been fed with whole cow's milk after the child is 1; recommending 2% cow's milk for children over 1 years old in the U.S. population. Another major milestone of the Bogalusa heart study is that identified several risk factors such as obesity, essential hypertension linked to kidney disease, and also how early onset of diabetes can also increase highly the development of cardiovascular disease at earlier ages, a finding that was also new to the medical and public health community.[28][29]

### The San Antonio Heart Study

The San Antonio Heart Study conducted in Texas takes its name from this city. San Antonio, Texas, is a city of the Southern United States, and it is considered one of the seven most populous city in the U.S. [30]



'San Antonio, Texas', image from Wikimedia, licensed CC BY-SA 4.0

The San Antonio Heart Study (SAHS), is another study that is no commonly mentioned in most epidemiology textbooks, and that brings another important perspective in the study of cardiovascular disease in the U.S. is the San Antonio Heart Study, which focused its efforts in identifying cardiovascular risk factors in the Latino population in the U.S. Again, as in the case of the Bogalusa Heart Study; the need to study in detail what happened

28. No author. (n.d.). Bogalusa Heart Study. From https://www.clersite.org/bagalusaheartstudy/
29. Berenson G. S. (2001). Bogalusa Heart Study: a long-term community study of a rural biracial (black/white) population. The American journal of the medical sciences, 322(5), 267–274.
30. San Antonio, Texas. Wikipedia. From https://en.wikipedia.org/wiki/San_Antonio

to another major ethnic group in the U.S. was critical, what was found among Caucasians or, Whites in the U.S. cannot necessarily be applied (extrapolated) to the African American community, nor, to the Latino community, so, the study was justified. And its findings among others discovered that the Latino heart is hard to die. Among all of the major ethnic groups in the U.S., Latinos have the lowest rates of heart attacks after accounting for several confounding factors.[31],[32],[33]

**Note:** the above content about cohorts had presented the problem of heart disease among the white, black and Latino population in the U.S. but there is literature available about other ethnic groups in the U.S. such as Native Americans, Asians, and Pacific Islanders, however, those studies are not cohort studies, only reports about the health status of these mentioned groups, for example, there is one report about Native Americans in the U.S.[34]

### Cohort Studies major disadvantages and historical 'mistakes'

Besides the disadvantage already mentioned before in the content, that cohorts are very expensive (they usually costs millions over the years of duration), it is not possible to have a cohort for every disease that exist, or, that is highly prevalence in the population. There is another limitation of cohorts, they cannot be used for the study of rare diseases, because one of the characteristics of a rare disease is that it is not suffered by a great number of people in the population, making the sample usually small, or, it may take long times to get enough data that can be used meaningfully in the practice.

Another major disadvantage of cohorts, which is the reason I included the examples of the 'famous cohorts in the U.S.,' is that in this country, cohorts were linked to the problem of racial preference, the first ethnic group represented in a popular cohort such as the Framingham heart study is the white population,

31. Shen D, Mitchell B, Hazuda H, Clark G, and Stern M. (1992). The San Antonio heart study research information study, Proceedings Computers in Cardiology, 607-610. Available at: https://ieeexplore.ieee.org/document/269385/keywords#keywords
32. Haffner, S. (2000). Obesity and the metabolic syndrome: The San Antonio Heart Study. British Journal of Nutrition, 83(S1), S67-S70. From https://www.cambridge.org/core/journals/british-journal-of-nutrition/article/obesity-and-the-metabolic-syndrome-the-san-antonio-heart-study/1E337C3FA2476672AD426FB90209EC10
33. Haffner SM, Miettinen H, Gaskill SP, Stern MP. (1996). Metabolic Precursors of Hypertension: The San Antonio Heart Study. Arch Intern Med. 156(17):1994–2001. doi:10.1001/archinte.1996.00440160106013 Access to abstract from https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/622459
34. Breathett, K., Sims, M., Gross, M., Jackson, EA., Jones, EJ et al. (2020). Cardiovascular Health in American Indians and Alaska Natives: A Scientific Statement from the American Heart Association. Circulation, 141:e948–e959. From https://www.ahajournals.org/doi/full/10.1161/CIR.0000000000000773

which excluded the other ethnic groups in the U.S. who also are affected by heart disease. This fact is an example of the historical exclusion of people of color and indigenous populations in the U.S., so, the history of major cohorts in the U.S. is also reflecting the need to study those oppressed, and ignored throughout history. Another observation in this context is, that for example, the Bogalusa Heart Study, and the San Antonio Heart Study are not well known in the scientific community, teaching medical schools, and other similar educational institutions do not know – or, pretend to not knowing about the existing of these studies, which provides extremely important data for the prevention of coronary heart disease in the nation.

*Clinical Trials*

They are considered the highest level of control of the study variables and interventions.

**Clinical Trials**

They are considered the highest level of the research designs discussed above in this chapter, and they are very much the standard design used especially for pharmaceutical companies to assess the effectiveness and safety of drugs, certain medical procedures, sophisticated medical equipment, etc. There at least *two types of clinical trials* as it was mentioned in the introduction of this chapter in the summary of types of study design; *1) Preventive or, Community Trials, and 2) Therapeutic clinical trials*. The discussion in this chapter will be mainly focused in the second group or, therapeutic clinical trials, also, just called, clinical trials.

**Community (or, preventive) trials**

These type of studies are used to determine the potential benefit of new policies and programs. They are called, **community** because it refers to the population, or, specific groups in the population. In general, the community trials will evaluate the impact of specific **interventions** that intent to produce changes in a target population. For example, the knowledge, attitudes and practices related to the Medicare program; or, the use (by the target population) of health care services to prevent and treat heart disease, etc.

The first step in the process of a **community trial** is to determine eligible communities, or, groups, and their willingness to participate. Then, baseline data is collected, this type of information can be for example, target population demographics, cultural traits, data from the national census, disease rates, etc., of the problem to be addressed in the intervention, and the collected information is also used for the control communities. In addition, the trial participants are selected by **randomization** (which is described in more details later in this section) and the selected individuals (or, groups) are followed over time. Data is entered, analyzed, and reports generated. Finally, the outcomes of interest are measured and used to assess the effectiveness, or, to identified weak points in the program intervention, and how to improve the quality of the programs and services offered to the target population, or, group. An example of a community trial and its protocol is summarized in the image below:

'Design of rural community-based trial part of the study'. CLBD: chronic low back disorder; NP: nurse practitioner; PT: physical therapist. Image from Research Gate.

### Advantages and disadvantages of community trials

The major advantages of community trials is that are unique in providing information that be can used to estimate the impact of change in the behavior or modifiable exposure of the incidence of disease in a community or, group; and also, the effectiveness of services and programs offered to the target group. As any other study design, the community trials have some also some disadvantages, for example, in general they are considered inferior to clinical (therapeutic) trials – discussed in detail in the rest of this chapter); and that is because selection of participants into the study, delivery of the intervention, and monitoring of the study outcomes are not as strict (or, rigorous) as it is in a clinical trial. Other disadvantages is that the study results are affected by some population dynamics, especially secular trends because of the mobility, or, changes in the target population. Also, it is hard to avoid the influence of non intervention forces surrounding the study population or, group.

### Clinical trials

Since the content above has been mostly about the clinical community (or, preventive) trials. The information that follows will focus mainly in the conduction of therapeutic clinical trials, commonly called just, 'clinical trials.' What are **clinical trials**? A common definition is that, clinical trials are planned experiments that assesses the efficacy of a treatment (or, medical procedure) in people. It is medical research involving people. In a clinical trial, the study outcomes in a treated group are compared with outcomes in an equivalent control group. Participants in both groups are enrolled, treated, and followed over the same time period. [35]

### Methods commonly used in clinical trials

There is a series of methods or, strategies used for clinical trials, and these are at the same time considered the major strengths of clinical trials. The most important are discussed in the following paragraphs.

### Study Protocol

The clinical trials protocol is usually an extensive and detailed manual that outline the major steps of the study, especially outlining what could happen in certain situations during the completion of the study. For example, what to do is the investigators deviate from the originally planned study assignments of the participants? How many deviations in the protocol would be allowed? It is customary that for example, no more than three deviations would be allowed during the duration of the trial. Also, the protocol include the data collection instruments, data input procedures, and analyses once the data is collected. The presence of a protocol is a valuable tool to assure that the clinical trial is conducted under the required academic and scientific rigor.

35. National Institute of Aging. (n.d.). What Are Clinical Trials and Studies? From https://www.nia.nih.gov/health/what-are-clinical-trials-and-studies

One of the major elements of the mentioned document is to plan ahead for any deviation of the study protocol. To prevent for this to happen, planned **crossovers** are part of the protocol. In this case, the study participant may server as his/her own control. And, when unplanned situations occurs for example a change of treatment is requested by a study participant; this change is called, an *unplanned crossover*, which could exist in for example situations in which the study participant request a change of treatment. It is recommended that no many unplanned crossovers occur during the duration of the trial, because it can compromise the study results, and the quality of the study in general.

### Selection of the study participants

One of the methods (and strengths) of clinical trials is the careful selection of study participants. For this purpose, research in clinical trials used what is known as **randomization,**, a statistical method to sort the possible study participants before they can participate in the study. Essentially, the randomization procedure allows to use random selection twice, for example, a person is assigned a number that will be randomly picked for the study participation, and then, if selected, the person (now, a study participant) will be randomly assigned to the drug, procedure, placebo or, no treatment branch of the trial.

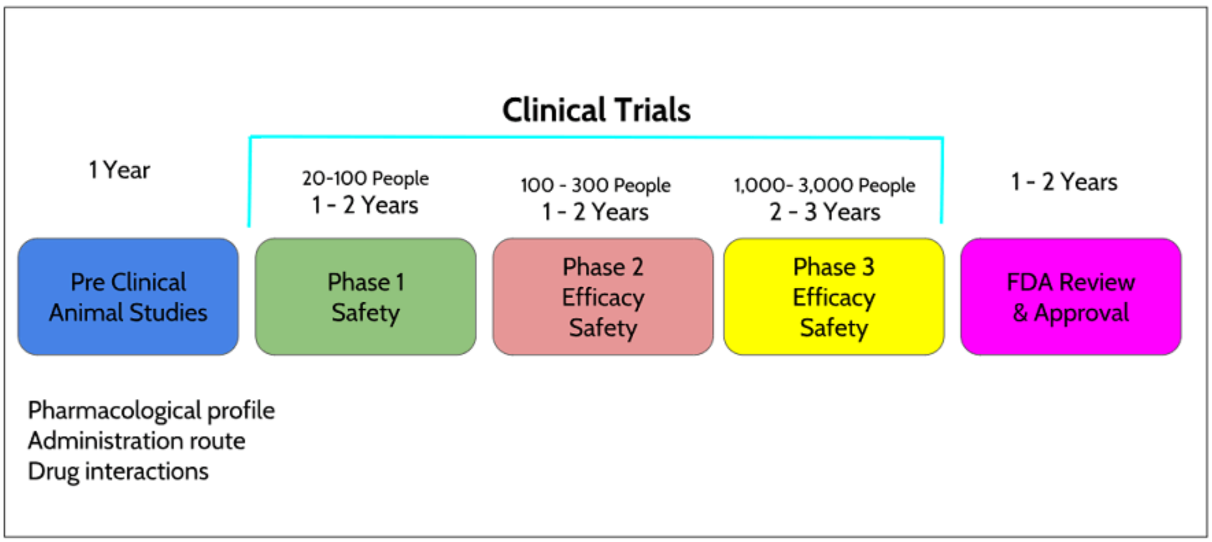
Randomization is the preferred method for assigning subjects to the treatment or control conditions of a clinical trial. If not random assignment is used then, mixing of effects of the intervention can occur, which at the same time, create differences among the study participants in the trial.

### Blinding

An additional method or strategy commonly used in clinical trials is also one of the major strengths of clinical trials is that they control bias, especially selection bias. And, to control for this, the procedure known as **blinding** is used. At least three types of blinding are known, 1) single blinding, it is when the study participants (clients, patients) don't know about the type of drug, or, medical procedure they have been assigned, or, they don't know if the treatment they are receiving is placebo. 2) double blinding, it is when the client/patient doesn't know what type of treatment or, placebo they are receiving (the clients/patient), or, the health care providers (doctors, nurses, technicians, etc.) are administering. 3) triple blind, the client/patient, the health care providers, nor the data collection and analysis have knowledge of the treatment given to the study participants. Additional blinding strategies can be used, but the mentioned here are the most commonly used.

### Phases of clinical trials

There are some established phases (or, steps) for clinical trials, they are used for example, before a vaccine, drug, or treatment can be licensed for general use. It is a lengthy process that helps to protect the public. The following image helps to illustrate the process:




'Clinical trials phases', image from JLI (James Lind Institute).

### Phases of clinical trials

These phases can be observed for example during the creation and testing of a new vaccine process follows:

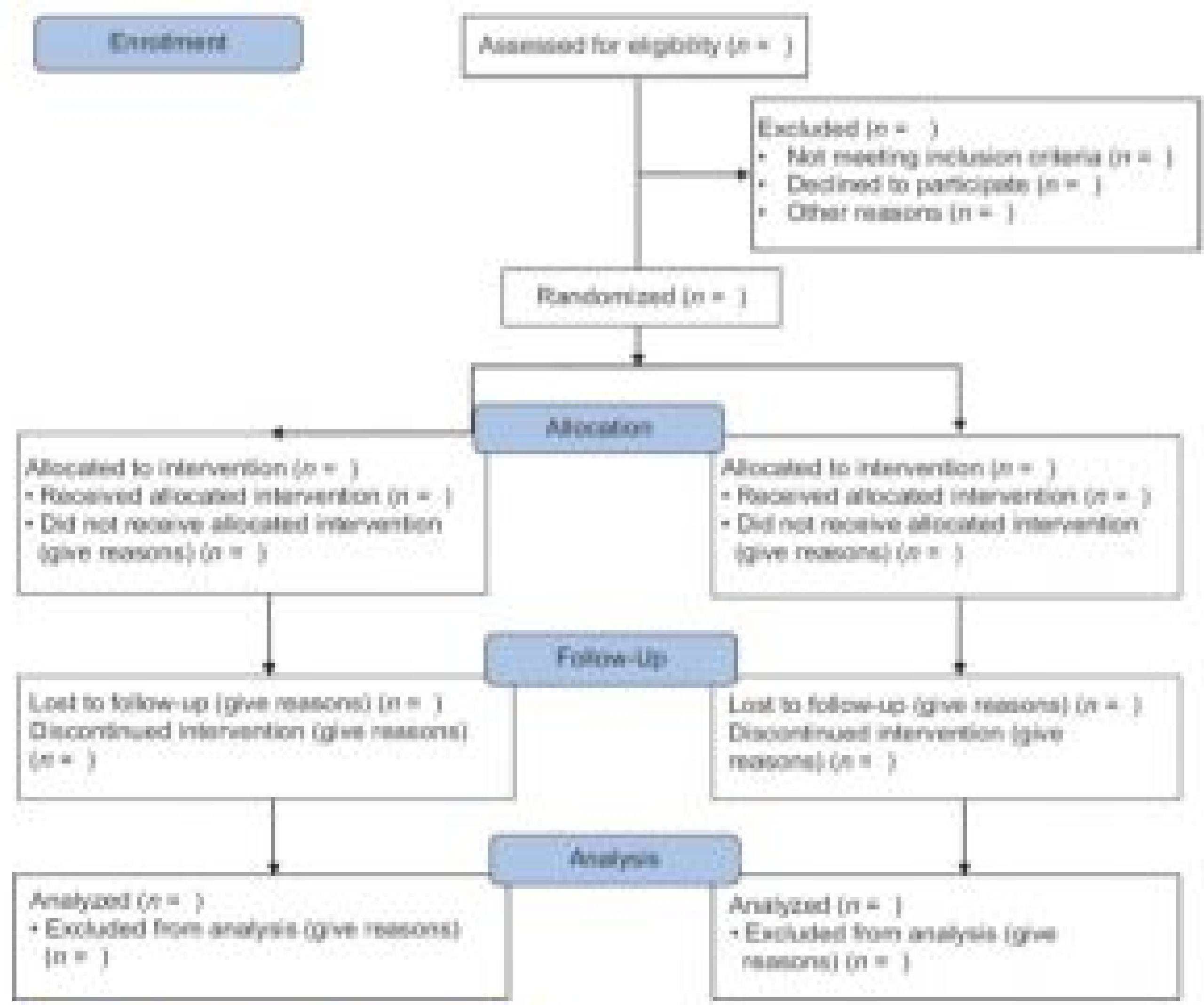
**Phase I**: tests a new vaccine in adult volunteers (usually with fewer than 100 volunteers).

**Phase II**: expands testing to a group of 100 to 200 subjects (from the targeted population, or group).

**Phase III** ( it is the main test), which intends to assesses the efficacy of the vaccine in the target population.

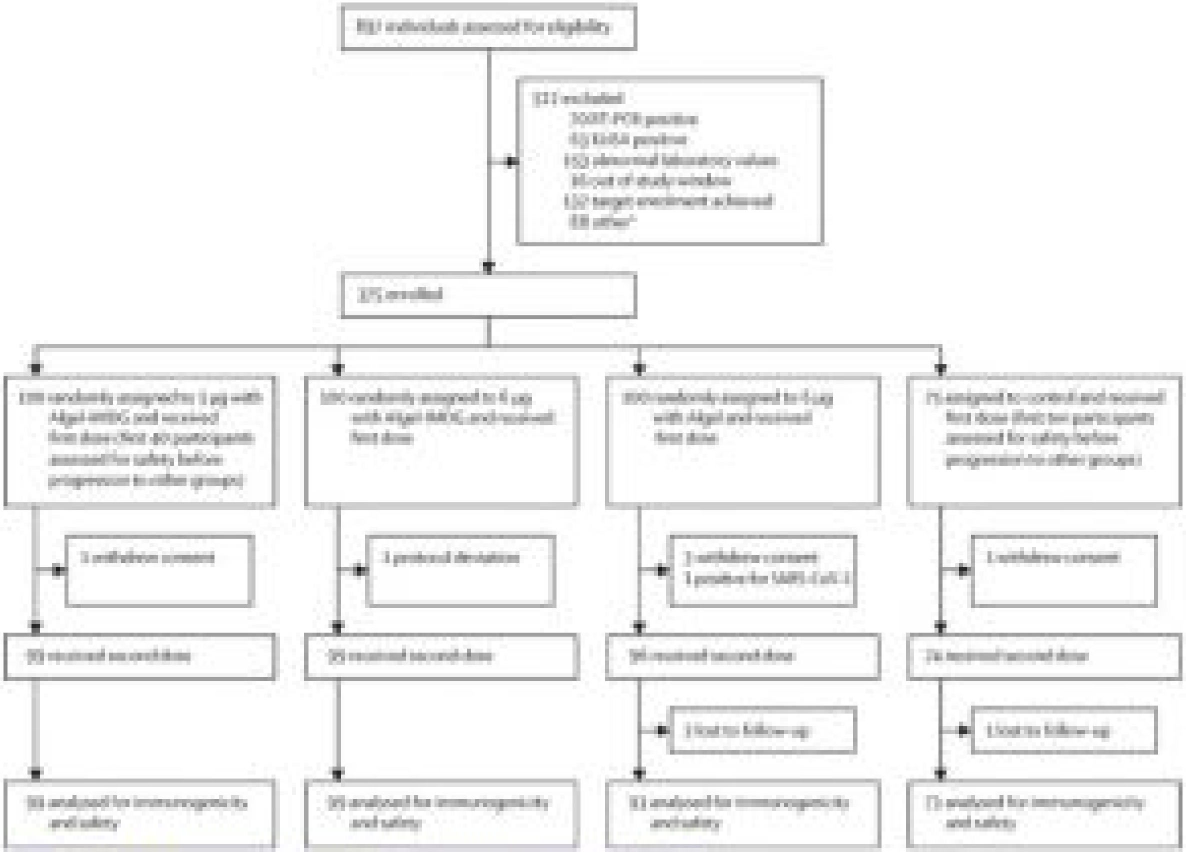
### Reporting of clinical trials – The Consort flowchart

There is a document called, the Consort Statement (a flowchart) that is a protocol used to guide the reporting of randomized trials by providing a 22-item checklist and a flowchart. The form is presented below:




"The CONSORT 2010 flowchart", taken from Pubmed

Example: During the COVID-19 pandemic, new vaccines have been developed and the following is an example of the use of the CONSORT flowchart used to describe the process. It is about a trial on severe acute respiratory syndrome coronavirus 2. See image below:



"Trial profile SARS-CoV-2=severe acute respiratory syndrome coronavirus 2". *Unable to contact the participant for vaccination or withdrawal of consent. Image from PubMed.

### Strengths and limitations of clinical trials

The *major strengths* of clinical trials is for the investigator to have the greatest control over the amount of exposure, the timing and the frequency of that exposure, and the observation period. Also, the use of randomization greatly reduces the likelihood that groups will differ significantly, which is of enormous benefit to the enhancement of the results. The *limitations* of clinical trials include ethical dilemmas such as how the benefits would outweigh the risk, how to protect the interests of the study participants and not only the investigators, when to stop a trial if a major adverse health outcome occurs during the completion of the study, and overall, how much of the information in the trial is shared with the study participant in the informed consent form.

### Summary

This chapter has covered the most common epidemiology study designs and its uses. These designs include: the case study/case reports, ecologic study, cross-sectional, case controls, cohort studies, and clinical trials. When they are classified by type of study, they can be descriptive or, analytic studies. The most common descriptive studies are, the case report/case series, ecologic and cross sectional studies. Analytic studies, they can be classified into intervention and observational. Examples of analytic 'intervention' studies are, community (preventive) trials and clinical (therapeutic) trials. And, observational studies, examples, the case control, and the cohort studies. The use of these mentioned study designs depending of the research questions, the purposes of the study, and the availability of resources to conduct them. Investigators always look for those study designs that are relatively quick, less expensive, and efficient.

# 6. Basic Epidemiological Methods and Calculations

*Learning Objectives*

By the end of this chapter, the learner will be able to

- Introduce the most common methods of disease frequency in epidemiology
- List common epidemiological methods of disease frequency: counts, proportions, ratios, rates, prevalence and incidence
- Apply the concepts of disease frequency in the analysis of results from epidemiological research studies
- Summarize the uses of prevalence, and incidence data in epidemiology.

### Epidemiology as a quantitative science

As it has been briefly mentioned in previous chapters, epidemiology is mainly a quantitative sciences. And, although this is changing in recent years in which more qualitative research is becoming more accepted, and available; the majority of professionals in the field accept that epidemiology is by nature, quantitative. So, what does it mean? It means that a quantity is used; epidemiology uses numbers (values), applied statistics, and other mathematical concepts to describe, analyze, and interpret health phenomena/events.[1]

### Prevalence, incidence, counts, proportions, ratios and rates

Common epidemiological quantitative techniques/methods include: counts, proportions, ratios, rates and also prevalence, and incidence.

1. Hernandez JBR, Kim PY. Epidemiology Morbidity And Mortality. (2020 Oct 13). In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan–. PMID: 31613448. Available at https://www.ncbi.nlm.nih.gov/books/NBK547668/Note: this above reference refers to the origins of the term 'epidemiology' and also epidemiology as a science but half of the article is also about common calculations in epidemiology.

“Woman who counts money,” By Jan Chalon.
Image from Look and Learn. Public Domain Dedication (CC0 1.0)

**Counts**

In some cases, and especially for the recording of data about some diseases (these include mainly rare diseases, or, those that had appeared in the twentieth century – such as the case of Ebola in the seventeens (Ebola virus was first described in 1976 near the Ebola River in what is now the Democratic Republic of the Congo) [2] In this case, all it can be done at that time is to record a count since the affected people represented only one small number, such as 1 case of ebola; . For example, in the past, we used ebola, which was very rare in the eighties; 1 case of ebola reported in 1976, with zero mortality at that time. [3]

**A** similar situation happened when the first cases of anthrax in the U.S. were reported, just after September 11 New York attacks. [4] When the initial publications

2. Centers for Disease Control and Prevention (CDC). (n.d. ). History of Ebola Virus Disease (EVD) Outbreaks, from https://www.cdc.gov/vhf/ebola/history/chronology.html
3. Emond RT, Evans B, Bowen ET, et al. (1977). A case of Ebola virus, British Medical Journal, 2(6086):541–544. From https://www.bmj.com/content/bmj/2/6086/541.full.pdf
4. Centers for Disease Control and Prevention.(n.d.). History of Anthrax, from https://www.cdc.gov/vhf/ebola/history/chronology.html

appeared, they referred only to 11 cases of anthrax.[5] As the time passes, and more cases are detected, and aggregated, the count may become higher, but it is usually not high enough to do more than just count the number of cases. No statistical analysis are needed in these mentioned situations.

### Ratios

When there is need to compare one number against another, in epidemiology, it is customary to use the ratio, which is a fraction (consisting of both numerator and denominator). A ratio is basically one quantity divided by another. There is no specified relationship between numerator and the denominator. The relationship is specified in the context of the problem that wants to be represented.[6]

> The formula for ratio is defined as **a:b ⇒ a/b**, where, **a** and **b** could be any two quantities.

The most common ratio is to compare the presence of a health problem among women compared to men. And, it is said, for example that breast cancer affect mostly women and just few men. As a quantitative expression, this can be said for example that comparing women to men, women have 100 times the possibility to have breast cancer compared to 1 men, expressed quantitatively, 100:1 or 100 to 1.[7] For this reason, it is said that ratios help to assess the impact of for example a health phenomena, commonly diseases as they are affecting people.

### Proportions

If more meaning wants to be added to a ratio, then, the proportion is used. Since, a proportion is a measure that states a count relatively to the size of the group; a proportion is basically a ratio in which the numerator is part of the denominator.

5. Brachman PS (2002). Bioterrorism: an update with a focus on anthrax. American Journal of Epidemiology, 155(11), 981-987. Also found at https://www.cdc.gov/anthrax/basics/anthrax-history.html
6. Centers for Disease Control and Prevention (CDC). (2012). Principles of Epidemiology in Public Health Practice, An Introduction to Applied Epidemiology and Biostatistics, Lesson 3 Section I. Third Edition. Available at: https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section1.html
7. Cancer Treatment Centers of America. (2019). What's the difference? Male breast cancer and female breast cancer. From https://www.cancercenter.com/community/blog/2019/07/whats-the-difference-female-male-breast-cancer

[8] That is the reason that it is said that, a ratio is a type of proportion that may be expressed as a percentage to help people better understand the meaning of the proportion, especially for health education purposes since most people are familiar with percentages but no with proportions.

The formula for a proportion is based on two ratios are a:b and c:d. Then, a : b :: c : d ⇒ a/b = c/d

Proportions are commonly used to describe the amount of disease that can be attributed to a particular exposure. For example, on the basis of studies of **smoking** and lung cancer, public health officials have estimated that greater than 90% of the lung cancer cases that occur are attributable to cigarette smoking.[9]

“Smoking Man,” By Jan Chalon. Image from Look and Learn. Public Domain Dedication (CC0 1.0)

8. Centers for Disease Control and Prevention (CDC). (2012). Principles of Epidemiology in Public Health Practice, An Introduction to Applied Epidemiology and Biostatistics, Lesson 3 Section I. Third Edition. Available at: https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section1.html
9. Centers for Disease Prevention and Control (CDC). (n.d.). Health Effects of Cigarette Smoking. From https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/index.htm

*Key Takeaways*

Caution: although the concepts so far defined in this chapter seem simple, don't get deceived, these concepts are not as simple as they seem,[10] especially when we used them in the practice.

**Rates**

If we put together the concepts about a ratio with the concept of a rate, it can be said that, a rate is a ratio that consists of a numerator, and a denominator in which, time forms part of the denominator. Adding the time dimension makes the concept of rate different from the rest of the concepts mentioned before. In addition, a rate contains the following: disease frequency, unit size of the population, and the recording of the time period in which the health event occurs. Examples of rates will be provided in detail later in this chapter.

**Moving beyond the basics**

The concepts of counts, ratios, proportions, and rates have its application in epidemiology as they are the foundation -concepts that help to understand two more important measures of disease frequency: *prevalence, and incidence.*

**Prevalence**

When there is need to measure the presence of disease/or, any other health phenomena, the most common measure that incorporates the time dimension (as it is the case of the definition of rate above) is, prevalence, which is defined a*s the number of existing cases of a health problem, or, health condition in a specific population at a certain (specific) time. The time dimension can be expressed in seconds, minutes, hours, days, etc.*

As we can see from the definition, there is need to understand the concept of case here (this concept will be explained more on detail as the content in this book progresses), but for practical reasons, *a case is a set of standard criteria that helps to identify those individuals/groups or, population who have been affected by a health*

10. Spronk, I., Korevaar, J.C., Poos, R. et al. (2019). Calculating incidence rates and prevalence proportions: not as simple as it seems. BMC Public Health 19,512. Available: https://rdcu.be/crLfC

*problem or, phenomena*. In general, the definition of a case can be obtained from the medical literature, or, from the local public health departments.[11]

**Prevalence** = number of existing cases/total population x multiplier and unit of analysis

Key:
Total population = or, study population
The Multiplier = it can be 10, 100, 1000, 100, 000 etc.
Unit of analysis = people, study subjects, experimentation animals, etc.

**Example of a hypothetical study:** In March 2010 a study on tuberculosis began, the target population is a recent group of immigrants located in the south side of St Cloud, MN. From this population, the study enrolled 500 individuals, and from this group, 50 residents were identified as confirmed cases of pulmonary tuberculosis. What is the prevalence of tuberculosis in this population?

**Prevalence** = 50/500 * 100 = 0.1 x 100 = 10 per 100 pop.

In some cases, people prefer to use the percentage to communicate the results, so, another way to report the results is to say that 10% of the study population had tuberculosis.

***Other elements added to this formula***

When the variable time is added to the formula (after the multiplier and unit of analysis), it helps to classify prevalence in two types: 1) **Point**, which refers to a point in time; and 2) **Period** prevalence or, the range of time is recorded and used in the calculation. When the data is sufficient in the calculation of the period prevalence, the variable time can be recorded/described in hours, days, weeks, months, years,

11. USAID. (n.d.). Basic Epidemiology. Primary Health Care Initiatives (PHCI) Project Contract No. 278-C-00-99-00059-00 Abt. Associates Inc. Available at: https://pdf.usaid.gov/pdf_docs/PNADN010.pdf

etc., depending on the type of health event being described. The formula for Point prevalence is presented below:

Point Prevalence = (# of existing cases/total population) x multiplier that includes the unit of analysis at a **point in time**

**Example:**

Following with the previous example and looking at the beginning of the study specifically on March 1st, 2020, the study investigators were able to identify 5 cases, which basically refer to individuals who already had tuberculosis before they were enrolled in the study. This information can be used to calculate the point prevalence (or a point in time), then, using the formula

**Point Prevalence** = 5/500 * 100 = 0.01 x 100 = 1 per 100 pop by March 1st, 2010 and if the information is expressed as a percentage, then, the result is 1%.

At the same time, if the duration of the investigation, or, the data collection period is known, the period prevalence can be calculated.

**Period Prevalence** = (# of existing cases/total population) x multiplier that includes the unit of analysis at a *period of time*

Using the same information example from above and knowing that, the study lasted 2 years, and now that this information is known, the result above can have an additional piece of information, which is the 2-years period in which a total of 50 individuals were identified with pulmonary tuberculosis. Using the formula above, the following is calculated:

**Period prevalence** = (50/500 *2) x 100 = (0.05 x 100) = 5 per 100 pop in a period of two years.

Or, (50/500) * 100 = 10/2 = 5 per 100 pop in a period of two years.

As seen in the result, the period prevalence added new meaning to the initial information provided by the simple prevalence calculation.

*Prevalence and incidence*

Prevalence and incidence are common measures of disease frequency used in epidemiology. Prevalence can be calculated in three ways: as a general measure, or, just as 'prevalence'; as a point in time, 'point prevalence,' or, during a specific period of time, 'Period Prevalence.'

**Prevalence common uses.** It is accepted that prevalence can be used to assess/evaluate the following characteristics of populations: to estimate the frequency of an exposure to for example an infectious agent, or, chronic problems such as cancer, or, cardiovascular disease. Also, to describe the burden of a health problem (which is the same as evaluating the impact of that health problem in the community); and determining allocation of health resources such as number of beds in a local hospital for specific health conditions; or, to identify the number of healthcare facilities, and health care personnel available to treat, or, prevent specific health conditions/problems.[12]

## Incidence

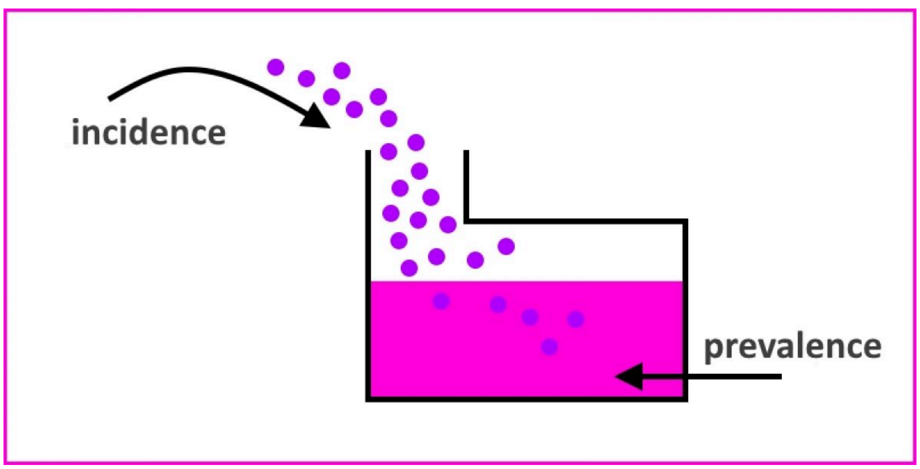
Another way to express the prevalence of a disease, or, health phenomena is to calculate its incidence. The formula for incidence can help define it:

12. No Author. (n.d.). Measures of disease frequency and disease burden available at: https://www.healthknowledge.org.uk/e-learning/epidemiology/practitioners/measures-disease-frequency-burden

**Incidence** = Number of New Cases/Population at risk x multiplier that includes the unit of analysis

*Key:*
*Population at risk = this is the segment of the population, who is affected by the health phenomena.*
*The Multiplier = it can be 10, 100, 1000, 100, 000 etc.*
*Unit of analysis = people, study subjects, experimentation animals, etc.*

In this manner, it has been also said that, that prevalence and incidence can be seen as the waterfall that feeds the pond. Recently, I've found another analogy which is illustrated below:



Incidence and prevalence Image from [Williamson, G.][13]

As can see in this image, the drops of water that feed the pool (are the new cases), and the water drops that fill the container are the total number of cases (new and old).

**To calculate the incidence** of the same hypothetical study used above, the following information is used:

The study that began in March 2010 lasted two years, with an initial number of

13. Williamson, G. (3 March 2016). Incidence and Prevalence. Categories: Research Methodology from https://www.sltinfo.com/incidence-and-prevalence/ Direct link to the image: Image from https://www.sltinfo.com/wp-content/uploads/2016/02/incidence-and-prevalence-1200x600-c-default.jpg

5 cases detected just at the beginning of the study, over the years, an additional number of 45 confirmed tuberculosis cases were found, these cases also belong to the total number of of 50 cases reported at the end of the two years. This information is used to calculate the incidence.

**Incidence** = Number of New Cases/Population at risk x multiplier that includes the unit of analysis

**Incidence** = (#new cases – #old cases)/population at risk * 100 = (45-5)/(50-5) x 100 = 40/45 x 100 = .88 x 100 = 88.8 per 100 of the pop or, 88.8% (if the result is expressed as a percent).

The result of the calculation shows that the incidence measures the progression of the disease, and the impact of the accumulation of cases in the calculation results.

Point prevalence, or, incidence are very useful to *quickly assess* the public health problem in a community, but for *health care planning*, for example how many beds are needed for a specific health problem, a period is most useful, because it helps to get an idea of the impact of the health problem on the community.

*Incidence calculation*

Incidence can be calculated as Incidence Rate (also known as Cumulative Incidence), Incidence for a period of time (known as Attack Rate – this will be cover almost at the end of class, and not for this unit), or, years of observation (this is called, Incidence Density).

**Additional example applying the concept of Prevalence and Incidence**

Suppose that in January 1996, 1000 adult residents of a community accepted an invitation to be examined for hypothyroidism at a local clinic. Eight persons were found to have the disease, it was newly discovered in 3, and 5 were already under treatment. The same group was examined again in January 1998. Six new cases of hypothyroidism were discovered; of these, two had developed symptoms several months before and had been diagnosed and treated by their personal physicians. It was learned that of the 8 hypothyroid persons discovered at the 1996 examination, one had discontinued medication and died of **myxedema** heart disease in 1997. Otherwise, all persons examined in 1996 came to the second examination. [The information provided in this example has been adapted from Friedman. Primer of Epidemiology.[14]. Also, the same example can be found in a packet of ancillary materials prepared by G. Antunez and available at Opendora Minnesota.[15]

Using the above information, respond to the following questions – don't forget to use the formulas already provided in this chapter:

1. What was the prevalence of hypothyroidism, treated or, not, in the examined group in January 1996? In January 1998?
2. What was the annual incidence of hypothyroidism in the group?
3. What was the 2-year period prevalence of hypothyroidism?
4. What was the case fatality rate of hypothyroidism?
5. Of all cases detected at the two examinations; what proportion was newly discovered?
6. If only 900 of the original 1000 persons were still living in the community and came to the examination in January 1998, would any of your answers to the questions above be changed? If so, how?Note: no answers for these questions are provided in the book. Please work on the responses, and consult with your class instructor if your answers are appropriate (or, correct).

**A Summary of the differences between prevalence and incidence**

In addition, if can be said that the main difference between prevalence and incidence is that incidence records, the number of new cases, and integrates the concept of population at risk. To help understand this concept, a comparison between incidence and prevalence is probably useful.

14. Friedman, G. D. (2004). Primer of Epidemiology. (5th ed). New York, NY: McGraw-Hill Companies.
15. Antunez, G. (2022). Principles of Epidemiology Ancillary Materials. St Cloud State University. From https://opendora.minnstate.edu/islandora/search/Antunez?type=dismax

**Table 1. General Differences between Prevalence and Incidence**

| | Prevalence | Incidence |
|---|---|---|
| Calculation incorporates in the numerator | All cases (old and new) | Only new cases |
| Calculation incorporates in the denominator | All population (every one) | Only population at risk |
| In terms of the variable time | It is usually a point in time | It incorporates periods, intervals, etc. |
| Reflects the event as | a snapshot of the disease/health event | a movie/video recording of the disease/health event |

*Note: Table created by the author using what it has been learned by years of teaching on the subject are, and consulting different sources.*[16]

### Incidence and the variable time

In the case of incidence, *time is essential*. Since we want to know the progression/ or, stopping of the 'new cases' during specific intervals/or, periods of time; it is possible to record, and report incidence in terms seconds, minutes, hours, days, weeks, months, years, decades, etc.; and this property of incidence makes the measure very useful in the control of disease, especially infectious diseases that depend a lot of the amount of time of its existence, or, presence in the community. But also, the variable time is important for the reporting of chronic diseases, for example, we want to know the survival rate of certain type of cancers; or, the number of disability years that mental health disorder can bring to the individual suffering from it.

**Incidence** = Number of New Cases/Population at risk x multiplier that includes the unit of analysis in a time period

Key:
The time period = it can be specific intervals/or, periods of time. For example, it is possible to record, and report incidence in terms seconds, minutes, hours, days, weeks, months, years, decades, etc.

16. Centers for Disease Prevention and Control (CDC). (n.d.) Lesson 3: Measures of Risk, Section 2: Morbidity Frequency Measures. Principles of Epidemiology in Public Health Practice, Third Edition. From https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section2.html

### Prevalence, incidence, and place/location

Besides the variable time, when is time to calculate prevalence, or, incidence, it is also useful to provide information about place/location of the health event. The most common places/locations used to report prevalence and incidence include: *country, region, continents, and other geographic locations*. However, we can use also *spatial locations* to report some health phenomena, or, events. Examples of these concepts are presented in the table below:

| Variable (type of place/location) | Example |
|---|---|
| Country | Prevalence of type 2 diabetes in the United S |
| Region | Prevalence of asthma attacks in school chilc |
| Continents | Prevalence of Malaria in Africa |
| Spatial | Prevalence of Depression in Urban Settings |

Note: to avoid confusion with the examples presented in this table, I only used the word, 'prevalence,' but the examples apply also to incidence. For example, the incidence of type 2 diabetes in the U.S.

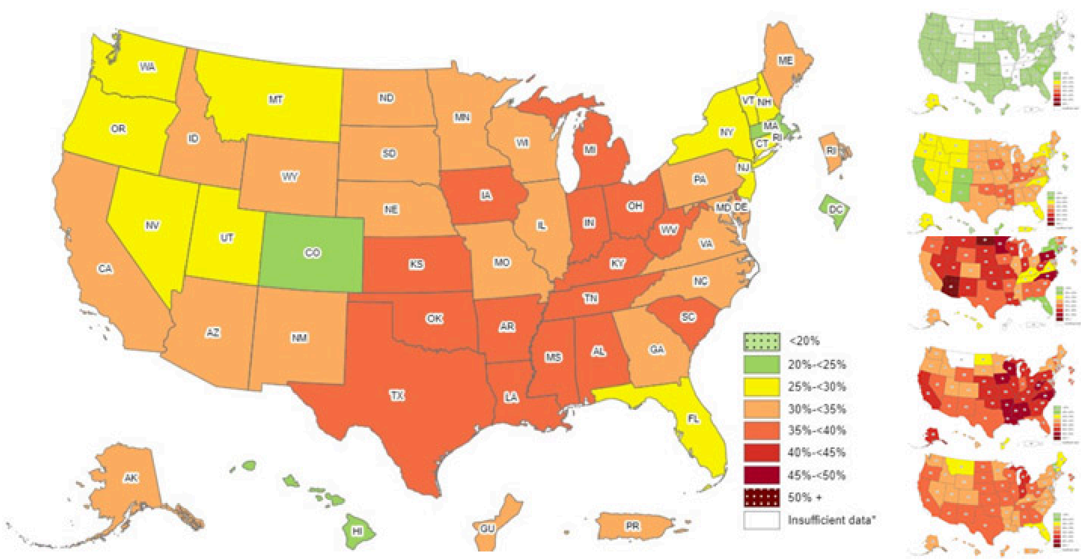
*Note: Table created by the author using what it has been learned by years of teaching on the subject are, and consulting different sources.*[17]

### When to add the time information?

For both (point prevalence or, point incidence), the place, or, location can be added depending of the amount of information that is requested, or, needed. A common combination include, point in time, and location. For example, prevalence/or,

17. Ford, G. (6th November 2020). Prevalence vs. Incidence: what is the difference? From https://s4be.cochrane.org/blog/2020/11/06/prevalence-vs-incidence-what-is-the-difference/#

incidence of obesity in the U.S. general population in 2019. In a most specific manner, the same information can be presented by adding specifics about age of the population, for example, prevalence/or, incidence of obesity among school children K-12 in the U.S. in 2019. Multiple combinations can exist by using place/location, age, gender, and other related factors. An image is presented below as an example of the prevalence of obesity among the adult population in the U.S. in 2020. The image is also part of the publication.[18]




[“Adult Obesity Prevalence Maps”.](https://www.cdc.gov/obesity/data/prevalence-maps.html) [Centers for Disease Prevention and Control (CDC)](https://www.cdc.gov/).

As the image shows, we can easily identify where the health problem of adult obesity is found at the state level in the country. Prevalence information is commonly used to develop this type of images/maps.

## Applications of Incidence Data

The concept of incidence has an important application, which is the calculation of **rates**.[19] Rates are used to express the incidence of disease in a population. Why

18. Centers for Disease Prevention and Control (CDC). (n.d.) Adult Obesity Prevalence Maps. From https://www.cdc.gov/obesity/data/prevalence-maps.html
19. No author. (n.d.). PH717 - Module 3 - Measuring Frequency and Association, Incidence. From

do not use prevalence instead? Prevalence is used to assess the extent of health phenomena mainly disease in a population, but it only allows to see an incomplete picture of the problem while incidence, especially expressed as rates is very useful to assess the impact or, extent in which the disease or health phenomena is affecting the population. It can also be used to assess possible **etiological** causes.[20]

*Additional resources*

The information presented in this section of the chapter can be also found for free online[21], and it should be good for consultation, or, to learn more about all of the concepts discussed in this chapter.

### Rates

Since they are based on the same formula of the calculation of incidence; the numerator reflects the number of specific cases, and the denominator captures the segment or total of a population who have been exposed to the health phenomena or, disease.[22] In this manner, the following is the **general formula** for a rate:

**Rate** = Total number of cases/Total Population (or, population at Risk if available) x multiplier that includes the unit of analysis

Key: Population at risk = this is the segment of the population, who is affected by the health

https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module3-Frequency-Association/PH717-Module3-Frequency-Association4.html

20. Ward M. M. (2013). Estimating disease prevalence and incidence using administrative data: some assembly required. The Journal of rheumatology, 40(8), 1241–1243. https://doi.org/10.3899/jrheum.130675
21. Centers for Disease Control and Prevention (CDC). (2012). Principles of Epidemiology in Public Health Practice, An Introduction to Applied Epidemiology and Biostatistics, Lesson 3 Section I. Third Edition. Available at: http://www.cdc.gov/csels/dsepd/ss1978/
22. Boston University, School of Public Health. (n.d.). Summary of Measures of Disease Frequency. From https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module3-Frequency-Association/PH717-Module3-Frequency-Association4.html#headingtaglink_7

phenomena.
The Multiplier = it can be 10, 100, 1000, 100, 000 etc.
Unit of analysis = people, study subjects, experimentation animals, etc.

*Types of Rates*

*Rates can be at least of three types: crude, specific, and adjusted. Crude rates are summary rates, and their calculation include the total population over a specific, or, a period of time.*

### Crude versus Specific Rates

When a rate is expressed as an analysis of the impact of disease/health phenomena in a population without taking into account much specifics, then, *the rate is referred as 'crude.'* It is crude because is not specific, it just point to the need to intervene, and to go into more detail in the exploration of the situation or health problem. That is the main reason, the *specific rates* are preferred if available. Specific rates are usually calculated using common variables such as age, gender, ethnicity, and other demographic characteristics that help to express the impact of these factors (variables) among the affected population. So if available, specific rates are preferred to crude rates. And, when the information needs to be more and more specific, there is need to adjust the information in order to eliminate the confounding factors, and the method to assess this is to use adjusted statistics, which gives the name to a third type of rates, the *adjusted rates.*[23] [24]

*Specific rates*

*Specific rates, refer to the*

### What are the most common rates used in epidemiology?

The most common use of rates is to study the impact of health phenomena (disease or not disease, healthy people and other characteristics of wellness) among populations and inside these population, there is need to find out who is affected, what gender, what age group, etc., is affected by a

23. No author. (n.d.). Standardized Rates of Disease, Crude Rates. From https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_standardizedrates/ep713_standardizedrates2.html
24. Naing N. N. (2000). Easy way to learn standardization : direct and indirect methods. The Malaysian journal of medical sciences : MJMS, 7(1), 10–15. From https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3406211/#:~:text=The%20difference%20between%20crude%20rates,as%20standard%20(Figure%201).

*calculation of specific characteristics of a population as for example, age, gender, ethnicity, and they usually refer to a specific segment of a population.*

health problem or disease, or who are the healthy individuals in common population groups. For example, maternal and child health programs want to know how many women are pregnant, and how many of these women are giving birth to healthy babies. **Rates** are used to report this information, and that is the reason for the calculation of rates among those women who for example completed their pregnancy and deliver healthy babies. In this context, some of the most common rates are those related to maternal and child health. Examples include, the birth rate, fertility rate, infant mortality rate, maternal mortality rate and others.[25] It is also important to note that for the content of this section, the main focus will be on Crude, and Specific Rates, and most of the time, mortality rates.

**Examples of common rates calculation**

For the following examples, most of the information is provided here, but in other cases, there is need to find the information by searching the National Census, the CDC, or, any other source of credible information.

**Example 1:** There were 1,986,000 deaths in the United States in 1982. **What additional information is required to compute the crude mortality rate for 1982?** To complete the calculation, here is the formula (which is a modified form of the one presented above).

**Crude Death Rate** (U.S. 1982) = Total number of deaths during the year of 1982/U.S. Total Population in 1982 x million people (also, 100,000 can be used as a multiplier).

Note: The information for the U.S. in 1982 can be found at the United States Census Bureau.[26]

25. Health Knowledge. (n.d.) Rates and ratios used to measure health status including geographical, occupational, socio-economic position and other socio-demographic variations. From https://www.healthknowledge.org.uk/public-health-textbook/health-information/3b-sickness-health/rates-ratios-measure-health
26. United States Census Bureau. (2017). Data. From https://www.census.gov/data.html

**Example 2:** The following is **hypothetical data**: A city has a population of 100,000 people (45,000 males and 55,000 females), and 1,000 people die per year (600 males, and 400 females). There were 50 cases (40 males, and 10 females) of lung cancer per year, of whom 45 died (36 males and 9 females). With this information complete the following calculations

a) Crude mortality rate
b) Sex-specific mortality rate
c) Cause-specific mortality rate for lung cancer
d) case fatality rate for lung cancer

<u>**Note:**</u> For each calculation above, a modified version of the general formula (provided in the example 1) needs to be created by the reader/student.

**Health Indicators**

There are a series of rates that are used to assess the overall health of a population, especially the group of children and women in a society. The results of these rates are used as 'indicators' of the health of a population or, group. [27]For example, is important to know how many children are born alive as it is also important to find out how many children are born death; and the results of these rate calculations are the 'indicators,' which 'indicates' the health of the overall population. The following are the most common on these indicators:

**Birth Rate**

As discussed before in this chapter, the birth rate can be expressed as crude or specific (less common in this category) being the most common, the crude birth rate, which is expressed in the following formula:

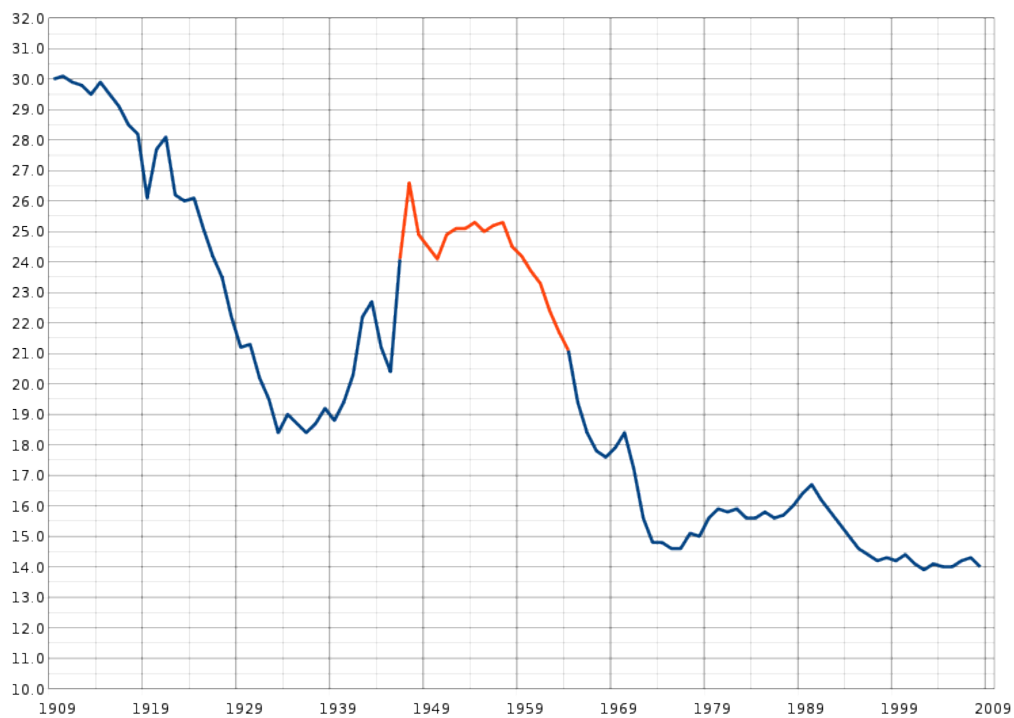
**Crude Birth rate** = # of live births within a given period/mid-population of that period x

27. Canadian Institute for Health Information. (n.d.). Health indicators. From https://www.cihi.ca/en/health-indicators

1,000 pop

*Note: in this formula to complete the calculation, it is customary to use 1, 000 as the multiplier. Please note that only the 'live births' are included in the numerator.  Also, in this case, the crude rate is the only 'necessary' calculation since the numerator is already 'specific,' only those babies that are born alive.*

In the following image, a decline in the overall birth rate has been observed progressively in the United States.



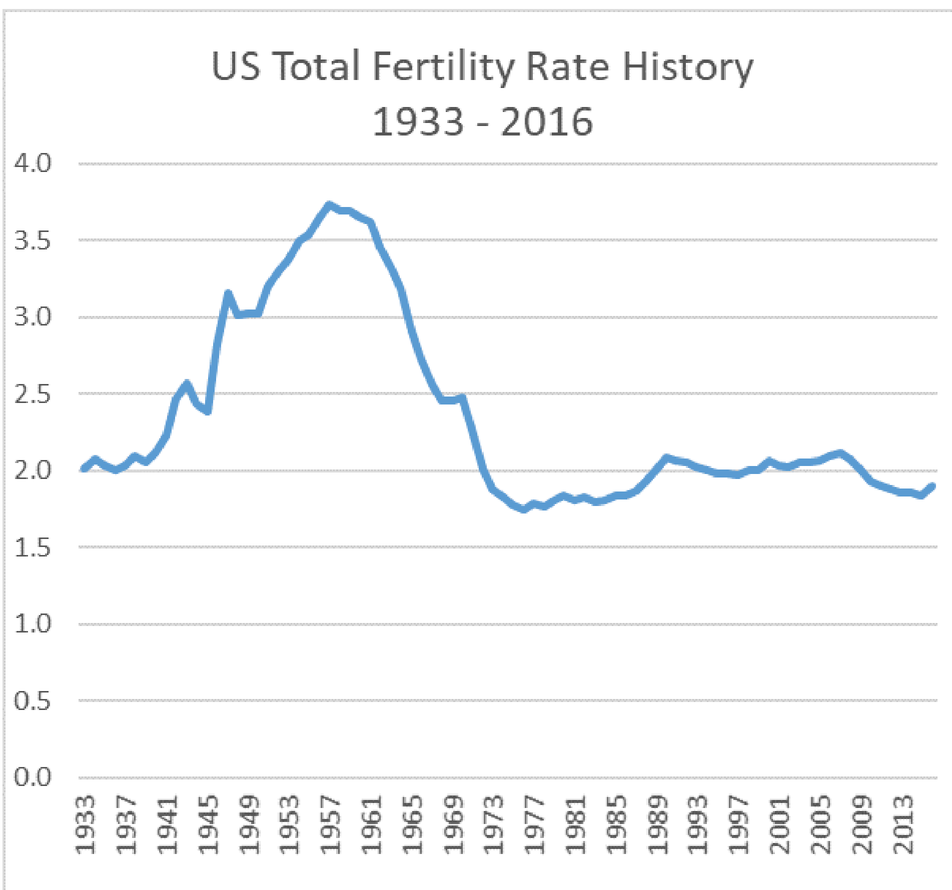

United States Birth Rates. No author. Source. Image licensed CC SA.

### Fertility Rate

This type of rate is commonly used by demographers, and organizations that work in the health of women because it is important to know how many women in the population are in reproductive age, which means able to become pregnant, which increases the population and the need for health services especially, reproductive health. The formula is:

**Fertility rate** = # of live births within a year/# of women age 15-44 years during the midpoint of the year x 1,000 women aged 15-44

*Note: It is important to note in this formula and in the denominator the specific target group is only those women who are considered (medically and for public health purposes) in reproductive age or aged 15-44.*




United States Fertility. No author. Source. Image licensed CC SA.

As seen in the graph/image above, the Fertility rate in the U.S. is in decline (In 2020, the U.S. TFR dropped to 1.64, the lowest level ever recorded), the same is happening in another major countries in the world. [28]

**Fetal Mortality Rate**

In the calculation of this rate, the numerator is represented by the numbers of children who died during the fetal period of life. In this case, there are at least two types of rates in this category, 'fetal death rate' itself, which is similar to 'early fetal death,' and the 'Late fetal death,' calculation formulas are included below:

**Fetal Death rate** (per 1,000 live births plus fetal deaths) = # of fetal deaths after 20 weeks or more of gestation/# of live births + number of fetal deaths after 20 weeks or more of gestation x 1,000

*Note: The term, 'fetal' is used in this formula instead of 'children' because medically a child that has not been born is a fetus. This may bring disagreement in the society, but appropriate or not, the term, 'fetus' is still the term in use in science. Also as it is accepted in medical sciences, a fetus that die when it is less or, equal to 20 weeks of gestation is considered a miscarriage (or, spontaneous abortion), or, abortion if it is not spontaneous. Finally, it is important to note (in the above formula that the denominator include the 'live births' and not only the 'fetal deaths'. The multiplier in this case is by convention the 1,000 and the result does not have to be labeled in the unit of analysis category.*[29]

Fetal mortality has been in decline in recent years in the U.S. as an article from an article published by U.S. News. [30]. The data on the article is based on data provided

28. Population Reference Bureau. (May 6, 2021). Why Is the U.S. Birth Rate Declining? From https://www.prb.org/resources/why-is-the-u-s-birth-rate-declining/
29. No author. (n.d.). Fetal Mortality Rate. From https://www-doh.state.nj.us/doh-shad/view/sharedstatic/FetalMortalityRate.pdf
30. Emerling, G. (2021). U.S. Fetal Mortality Rate Hits Historic Low The measure marks what federal researchers call 'a major but often overlooked public health issue.' U.S. News. From https://www.usnews.com/news/health-news/articles/2021-10-26/fetal-mortality-rate-hits-record-low-in-us

by CDC, which include images and tables not shown here for reasons of system's compatibility.[31]

### *Late Fetal Death Rate*

**Late Fetal Death rate** (per 1,000 live births plus late fetal deaths) = # of fetal deaths after 28 weeks or more of gestation/# of live births + number of fetal deaths after 28 weeks or more of gestation x 1,000

*Note: The emphasis here is in the word, 'late' which is a fetal death that occurs after 28 weeks of gestation. Again, the multiplier in this case is by convention the 1,000 and the result does not have to be labeled in the unit of analysis category.*

As a complement to the rates calculation, it is also customary to calculate the ratio of the 'fetal death' category. Again, in this case it is a ratio (not a rate), the formula is presented below:

31. Gregory, CW., Valenzuela C., Hoyert, D. (October 26, 2021). Fetal Mortality: United States, 2019 Division of Vital Statistics, National Vital Statistics Reports, 70 (11), 1-20. From https://www.cdc.gov/nchs/data/nvsr/nvsr70/nvsr70-11.pdf

**Fetal Death ratio** = # of fetal deaths after 20 weeks or more of gestation/# of live births x 1,000 (during a year)

*Note: As noted above, in this case (ratio calculation), the denominator does not include the fetal deaths as it is the case of the fetal death rate.*

### Infant Mortality Rate

The infant mortality rate is highly recognized in public health and that is because it is an important measure (that reflects) the overall health of a population since for every child that dies, there is also a mother who lost her child, and the reasons for the loss are reflected but the quality of care that a pregnant women receives in a specific society. There are people who believe that just by knowing the infant mortality rate (and the mortality rate too, which is presented later), a person can judge (or, evaluate) the overall health care and social system of a country. To calculate the infant mortality, the following formula is used:

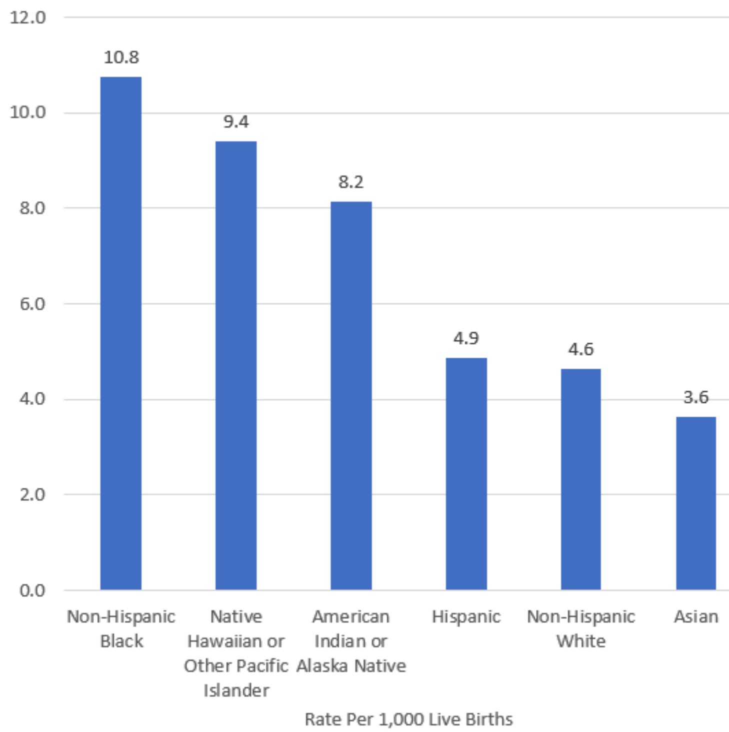
**Infant Mortality Rate** = #number of infant deaths among infants aged 0-365 days during the year/# of live births during the year x 1,000 live births

Note: It is customary (commonly used) that the multiplier for the infant mortality rate is set to 1,000 live births.

The CDC reported in 2019 that the infant mortality rate in the United States was 5.6 deaths per 1,000 live births, which represent 558.3 infant deaths per 100,000 live births, which did not change significantly from the rate in 2018.[32]

32. Kochanek KD, Xu JQ, Arias E. (2020). Mortality in the United States, 2019. NCHS Data Brief, 395. Hyattsville, MD: National Center for Health Statistics. From https://www.cdc.gov/nchs/products/databriefs/db395.htm

The state of Minnesota has a lower infant mortality rate, or, 4.47 deaths per 1,000 live births[33], which reflects that the state of Minnesota cares a lot compared most states in the U.S.




Infant Mortality Rate Ethnicity by Race 2017. No author. Source. Image licensed CC SA.

### Going deeper in the infant mortality indicator

For medical and public health interventions, the calculation of additional rates in different stages of the live of children is also considered important and useful. The following calculations related to this category follows:

33. Centers for Disease Prevention and Control (CDC). (n.d.) Infant Mortality Rates by State. From https://www.cdc.gov/nchs/pressroom/sosmap/infant_mortality_rates/infant_mortality.htm

**Neonatal Mortality Rate** = #number of infant deaths under 28 days of age/# of live births x 1,000 live births (during the year)

Note: It is customary (commonly used) that the multiplier for this type of calculation is set to 1,000 live births.

### *Postneonatal Mortality Rate*

**Postneonatal Mortality Rate** = #number of infant deaths from 28 to 365 days after birth/# of live births minus (-) neonatal deaths x 1,000 live births (during the year)

Note: It is important to note that the neonatal number of deaths is required before the postneonatal mortality rate is calculated.

## Perinatal Mortality

There is also two more calculations that are related to this category and that are used to evaluate the health of children during gestation and the first month of life. For these purposes, the perinatal mortality rate and ratio are used. To work on these this calculations the following formulas are used:

**Perinatal Mortality Rate** = #number of late fetal deaths after 28 weeks or more of gestation + infant deaths within 7 days of birth/# of live births + number of fetal deaths x 1,000 live births and fetal deaths

Note: It is important to note that the above is a rate and below is the ratio. Also, in the result the label used for the unit of analysis is both, live births and fetal deaths.

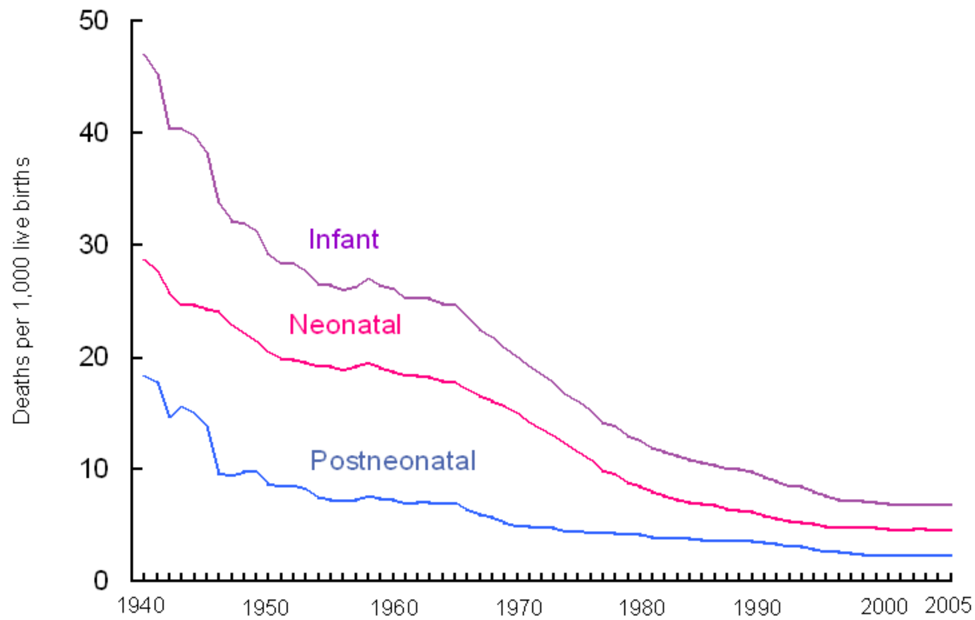
### *Perinatal Mortality Ratio*

**Perinatal Mortality Ratio =** #number of late fetal deaths after 28 weeks or more of gestation + infant deaths within 7 days of birth/# of live births x 1,000 live births

Note: It is important to note that the ratio calculation the denominator contains only the live births (excluding the fetal deaths).

The calculation of the perinatal mortality rate and ratio exists because in many cases, and in most countries of the world, there is insufficient or, no data that records late fetal deaths, making difficult if not impossible to calculate the rate, making the ratio calculation the most practical.

The author could not find an image that could be used for each one of the categories discussed above, but there is graph/image that gives an idea of the frequency of these types of deaths. See below:




United States Neonatal death. No author. [Source](). Image licensed [CC SA]().

Note: although the title of the image is not included in the graph presented here, the most appropriate title could be something as: "Deaths related to childbirth in the U.S. – A comparison).

### The health of the mother

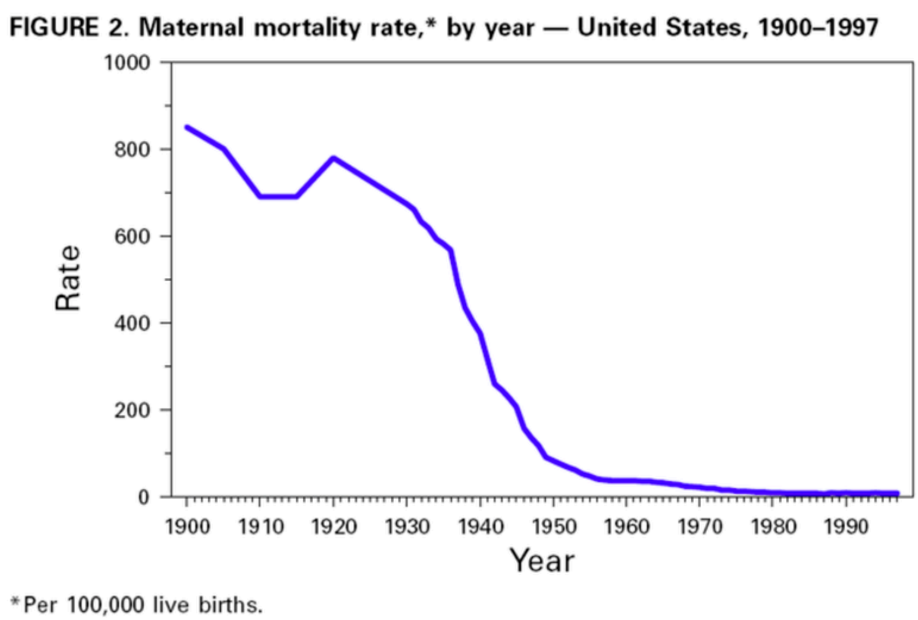
As it is important to assess the health of a child – calculations have been presented above, it is also important/crucial to assess the health of the mother, especially when there is a death of a woman due to or, related to childbirth. The formula for this calculation is presented below:

**Maternal Mortality Rate** (per 100,000 live births, including multiple births) = # of deaths assigned to causes related to childbirth/# of live births x 100,000 live births (during a year)

Note: It is customary that the multiplier for the maternal mortality calculation is 100,000.

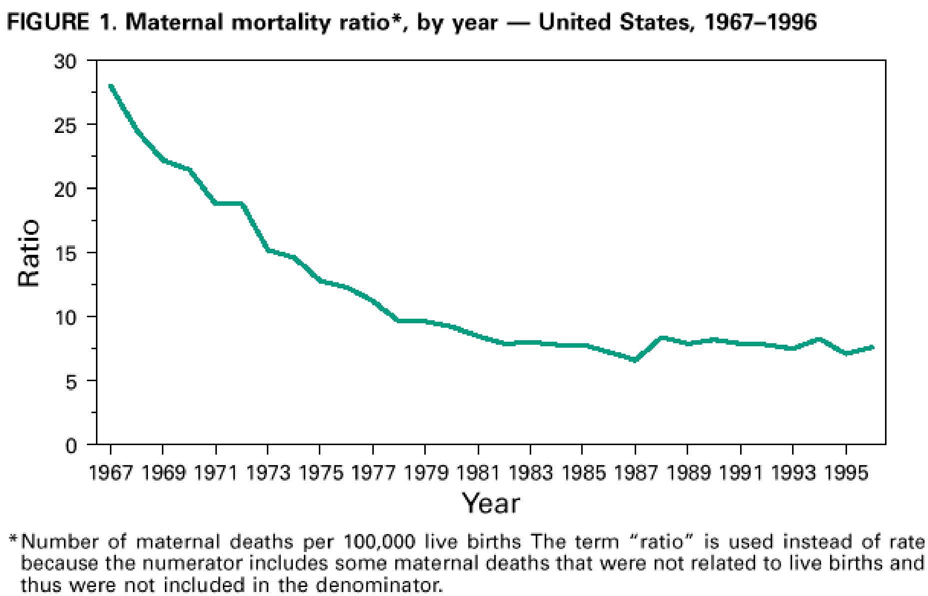
Maternal mortality is a strong health indicator of the health care conditions of a country, especially in issues related to pregnancy and childbirth. The more dissaraged a system is, the maternal mortality rates are high. It also reflects how poverty levels and distribution of wealth and care of women in a country or, region in the world. Based on this, it is accepted (although it is not always the case) that developed countries have lower indicators of maternal mortality (and also infant mortality) compared to developing countries.

Presenting visual examples of the problem of maternal mortality is always useful, and it is the case for example of the United States, which has seen a dramatic decline in maternal mortality from the 1900s to 1990, a graph with this information is presented below:




Maternal Mortality in the United States from the 1900s-1990. No author. Source. Image licensed CC SA.

The dramatic decline is not as dramatic when more current data is researched, here is one example:



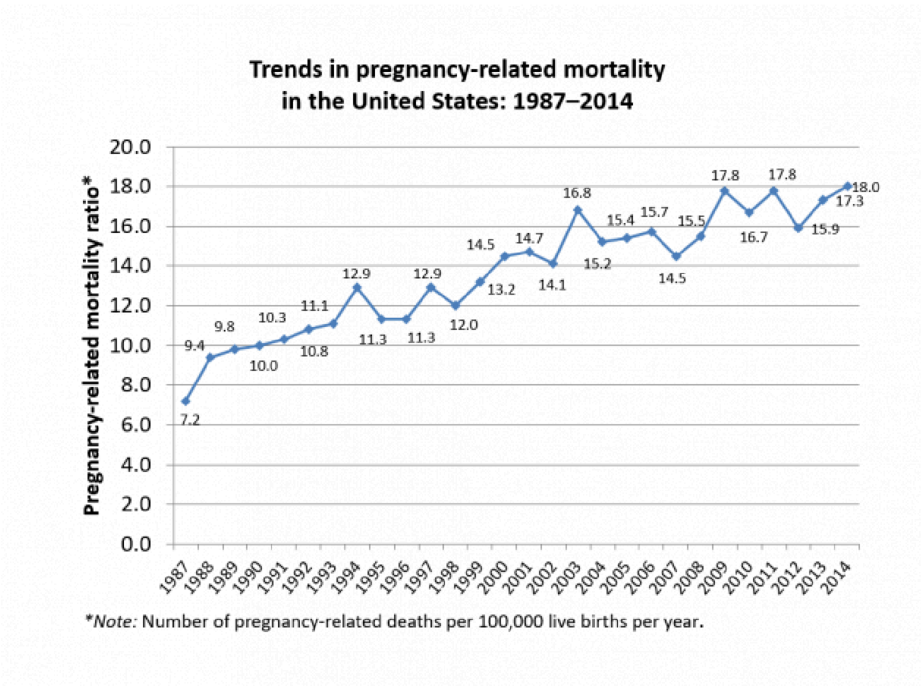

Maternal Mortality By year. No author. Source. Image licensed CC SA.

Although, the authors of this graph used the ratio instead of the rate, the information is good enough to make the following point: now, looking at these data, it is clear that the decline that seen so dramatic, it is not. In this publication the maternal mortality for that time period (1967-1996) was also reported as to be 3.3 maternal deaths per 100,000 live births. [34] Looking at the numbers on maternal mortality clearly show that the United States needs to make investments to improve the health of women, so, the maternal mortality is in reality declined. These mentioned investments should include above all, better health care system, with improved access to those women living under poverty, and who also are excluded from society such as for example homeless women.

A proof of the need to improve the health of women in general (including the decline of maternal mortality) is part of the healthy people 2030, which uses the current statistics, the United States Maternal mortality is 17.4 maternal deaths per 100,000 live births (2018), and this is used by the HP 2030 committee as the baseline data, with a desired target of reducing this indicator by 15.7 per 100,000 by the year 2030. [35] Over the years, I had a class activity/computer lab that asked students to find

34. Centers for Disease Prevention and Control (CDC). (n.d.). Maternal Mortality -- United States, 1982-1996. From https://www.cdc.gov/mmwr/preview/mmwrhtml/00054602.htm
35. Healthy People 2030. (n.d.). Reduce maternal deaths — MICH-04, LHI (Leading Health Indicators). From

the current data on maternal mortality and it has been always difficult, which for me, it shows that even data access about maternal mortality needs to be improved.




Trends in pregnancy related mortality in the United States. No author. Source. Image licensed CC SA.

**Practice calculations: Examples of some mortality indicators**

Using data from the National Census, the CDC, or, any other source of information that is accepted as credible calculate the most common type of mortality indicators:

**a)** Infant mortality in Minnesota in the year 2005 (use the formula provided before in this textbook, and search for the numbers needed to complete the calculation**.**

**b)** Neonatal mortality in the U.S. in the year 2005 (again, use the formula provided in the textbook, and find the numbers to do your calculation).

**c)** Maternal mortality in the U.S. in 2005.

https://health.gov/healthypeople/objectives-and-data/browse-objectives/pregnancy-and-childbirth/reduce-maternal-deaths-mich-04

Note:

1. To find the numbers to complete the calculations, search the United States Census Bureau.[36].

2. For all of the calculations above, be sure to always state the unit of analysis, for example, for infant mortality and neonatal mortality use the 1,000 live births, for maternal mortality, 100,000 live births (during a year).

### Specific Rates

In this category the topic of specific rates is covered with the addition also of the proportional mortality ratio. Both categories of calculations are useful to be more specific about for example, the 'specific' characteristics of the general population, or, a population segment or, study population. The formulas that follow cover these topics:

### Cause-Specific Rate

The cause-specific rate is useful to assess the impact that specific diseases or, any other health phenomena has a specific population group. The formula follows:

**Cause-Specific Rate** =Mortality (or frequency of a given disease)/Population size at midpoint of time period x 100,000

Note: the term 'midpoint' refers to for example, mid-year, or at the end of June. It has been found that statistically, most populations double their size in the second six months of the year (end of december). For this formula then, if the total population is known, the midpoint is found by dividing the number by 2. The use of the total population for the calculation is also acceptable.

As it is common to find the cause-specific rate, the age-specific rate is also common. The formula is:

36. United States Census Bureau. (2022). Data Census. From https://www.census.gov/

**Age-Specific Rate** = # of deaths among a specific segment of a population/# of persons who are in the same segment or group (during time period) x 100,000

Note: the term 'segment' refers to the group of people included in the calculation based on their age. Age frequency intervals are used to specify the age group. For example, the people (in the study population) whose age is 5-10 or 11-15, etc.

In addition, there is one more calculation in this category of calculations covered in this chapter, and it is the **Proportional Mortality Ratio**. As the name implies, it is a ratio (not a rate). The formula is:

**Proportional Mortality Ratio or PMR (%) =** Mortality due to a specific cause during a time period/Mortality due to all causes during the same period x 100

Note: the PMR multiplier it is by convention 100, which means that is a percentage (this is different compared with the rest of the calculations presented in this chapter). Also it is important to note that the mortality mentioned in the numerator has to be the same population included in the denominator.

Among other things, the PMR is not a measure of the risk of dying of a particular cause, it is in reality and indicator (in the study population) of the relative importance of a specific cause of death. It is commonly used to create the public health report known as, 'the 10 major causes of death in a population, or, country.' In general 10 causes are reported, but it can be more, for example 12 causes of death. It all depends of the main reason of the publication of this information, which is usually important for policy makers, and health care providers among others.

**Adjusted Rates.** The topic of adjusted rates is not be covered in this textbook, but the reader/learner can find more about it in the majority of advanced epidemiology textbooks.

**For additional reading on the concepts discussed in this chapter**

An overview of the concepts discussed in this chapter can be found online for free. Use this for consultation.[37]

**Summary**

This chapter has covered what it commonly called, basic epidemiological methods, which although seem to be simple compared to more advanced epidemiologic calculations methods, they represent the foundations of more complicated calculations. The chapter covers counts, ratios, proportions and then, the concepts of prevalence and incidence. Examples for all of these calculations have been provided in this chapter. It is also recommended that the student use the provided examples to expand on the concept, and practice more examples from the main sources of reference contained in this chapter. One great resource is free and it can be found online at the book, Epidemiology Kept Simple, whose link is provided above this summary.

37. Formulas from Epidemiology Kept Simple. (n.d.). Chapter 3: Epidemiologic Measures. Epidemiology Kept Simple. Third Edition. Available at: https://www.sjsu.edu/faculty/gerstman/eks/formula_sheet.pdf

# 7. More advanced calculations: Odds Ratios and Relative Risk

*Learning Objectives*

By the end of this chapter, the learner will be able to

- Introduce the most common methods to assess risk in epidemiology: the odds ratio, and the relative risk.
- Learn how to analyze and report the odds ratio, and the relative risks as common methods of assessing risk from the results of epidemiological research studies
- Summarize the uses of odds ratio and relative risk in epidemiology.

**Introduction**

In this chapter, the topic of risk assessment and its related calculations is discussed. The previous chapter on study designs has been intentionally covered before risk is assessed, and that is because each study design has a more favorable statistical method or calculation that fits better and as a consequence answers the main research question proposed by the investigator. The point is that the

concepts of study designs, and risk assessment in epidemiology are intrinsically connected. For example, for case study designs, the odds ratio calculation works better compared with the relative risk that is more appropriate for the cohort study design. However, both measures of association can be used for more than one study, which is also a matter of which measure of association calculation responds better to the need of the study, and the expectations of the investigator. Even in the case of clinical trials that used commonly for therapeutic reasons, the major common measures of association discussed in this chapter are also useful. So, the principles are the same, use a risk assessment method that better fits your needs.[1]

1. Knol M, J, Algra A, Groenwold R, H, H. (2012). How to Deal with Measures of Association: A Short Guide for the Clinician. Cerebrovasc Dis 33:98-103. Available at: https://www.karger.com/Article/Fulltext/334180#

**Risk assessment – the basics: Odds Ratios and Relative Risk**

One of the main reasons, epidemiology is so popular in public health, and medicine, is that it provides a platform to conduct research studies used to assess risk. A clinician may want to know what is the risk for the patient to die prematurely, or, long the patient has to live if there is a terminal disease such as cancer. The response for these questions come from epidemiological studies that had assessed the risk. The same is true for the public health practitioner, when there is a question about the risk of infectious diseases for an unvaccinated population, or, the access to health services for an underrepresented group in the community. Clinicians, and public health practitioners, both rely in epidemiology to answer their questions, and to plan health services among other interventions.


At Risk Assessment by Nick Youngson CC BY-SA 3.0 Pix4free.org

**But overall, what is risk, and what is risk assessment?**

In mathematical terms, the definition of***risk*** is, the probably that something will occur (happen) in a certain period of time.[2] It is accepted that in order to be

2. Online Dictionary (n.d.) Definitions from Oxford Languages. From https://www.google.com/

considered valid, this probability has to be greater than one, and less than zero [3] (more about this concept will be discussed later in this chapter). If we broadly had defined the concept of risk, what is risk assessment? ***Risk assessment*** is a series of techniques used to assess, or, evaluate risk. [4] It is based on the probabilities (risk), but it also includes other factors that contribute for the event to happen, these factors are also considered ***risk factors*****,** which are factors associated with the health event (disease, health, etc.), but that are not necessarily the cause of the disease. [5] The cause of the disease is usually assessed using other epidemiological factors that will be also discussed later when referring to the applications of research designs in epidemiology.

Image by
Wokandapix from Pixabay

**The concept of risk and how it is measured**

There are two major mathematical (or, statistical) methods used to assess (or, quantify) risk. The ***odds ratio***, and the ***relative risk***. Both of these methods required

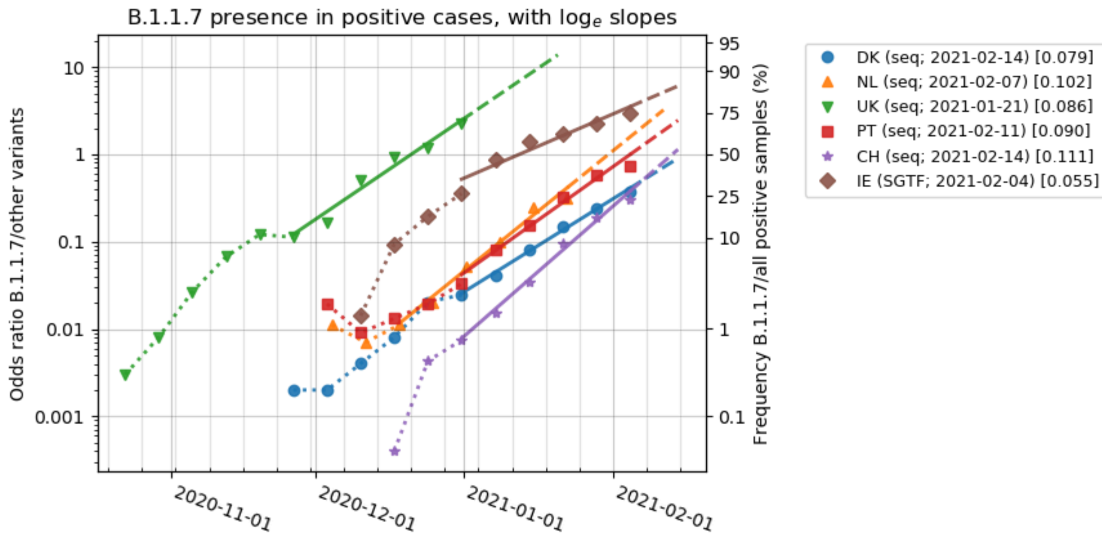
search?client=firefox-b-1-d&sxsrf=APq-WBueRS_Jgug2a61MY29uKUB-p22K9A:1647222052957&q=define+probability&forcedict=probability&dictcorpus=en-US&expnd=1

3. BYJU'S. (n.d.). Probability Definition in Math. From https://byjus.com/maths/probability/
4. Canadian Centre for Occupational Health and Safety (CCOHS). (n.d.). Risk Assessment. From https://www.ccohs.ca/oshanswers/hsprograms/risk_assessment.html
5. RxList. (n.d.). Definition of Risk Factor. From https://www.rxlist.com/risk_factor/definition.htm

calculations based on the probably of the events to happen. The results are commonly called, ***measures of association***.[6] In this manner, the results of the odds, or, the relative risk are used to assess the possible association of a risk factor in a health event (disease, or, not).[7] For this reason, in epidemiology, *the most common measures of association are the odds ratio, and the relative risk*.

And example can be seen during the COVID-19 pandemic as the following image presents, the odds ratios have been presented in a visual form, so, the reader interested in this information can find it easily. See image below:

6. Haug, M. Gerard (n.d.). measure of association. Encyclopedia Britannica. https://www.britannica.com/topic/measure-of-association
7. Alexander LK, Lopes B, Ricchetti-Masterson K, Yeatts KB. (2012). Common Measures and Statistics in Epidemiological Literature In Eric Notebook, Second Ed. Available at: https://sph.unc.edu/wp-content/uploads/sites/112/2015/07/nciph_ERIC3.pdf




"COVID-19 Odds ratios by country" Image by Hankwang, Licensed CC BY-SA 4.0 International, Wikimedia Commons. The image used here is from 2021-01-31, last updated 2021-02-14

Note: The following comments have been modified from the image website: As seen above in the graph, the odds ratio (left y axis) of B.1.1.7 versus 'wild' variants is [was] expected to grow exponentially; [0.076] means 7.9% growth per day. Also, as the author said: "The logarithmic slopes as indicated are not based on a rigorous uncertainty analysis."[8]

## Defining the concepts of Odds Ratio and Relative Risk

In order to understand these concepts (and also, for the calculation of these measures of association), there is need to review the concept of 2×2 tables first. Why? ***The 2×2 table assists with the visualization of the risk probabilities***. A 2×2 table is called that way because it has two rows (horizontal), and two columns (vertical). By convention, the cells are filled with letters, which later will be used to create the calculation formulas.[9] A 2×2 table looks something as the following:

8. Hankwang. (202). COVID-19 Odds ratios by country. From https://commons.wikimedia.org/wiki/File:Covid-19_VOC-202012-01_odds_ratio_curves_by_country.png
9. Sauerbrei, W., & Blettner, M. (2009). Interpreting results in 2 x 2 tables: part 9 of a series on evaluation of scientific publications. Deutsches Arzteblatt international, 106(48), 795–800. https://doi.org/10.3238/arztebl.2009.0795

**Table 1. The 2×2 table**

| | |
|---|---|
| a | b |
| c | d |

Later in this chapter, when the calculation formulas are presented, the totals (from the rows, and the columns) will also be visualized.


Odds by Nick Youngson CC BY-SA 3.0 Alpha Stock Images

**Odds ratio (commonly abbreviated as OR)**

In general terms, ***the odds is the probability that an event will happen divided by the probability that it won't happen.)*** One way to see (explain, visualize) this is to label, the 2×2 cells in terms of exposure, and not exposed. The layout of the table will look something as this:

**Table 2. The 2×2 table with labels in the cell categories**

| Exposed | Non-exposed | Totals |
|---|---|---|
| a | b | a+b |
| c | d | c+d |
| a+c | b+d | a+b+c+d |

Then, it can be seen that *the odds ratio is a measure (calculation of the probabilities) that compares the odds of exposure to the factor of interest among* *cases to the odds of exposure to the factor among controls*. That is one of the main reasons, the odds ratio is the most commonly used calculation to assess risk in case-control studies, in which cases are compared to those that are non-cases (controls, or, those not affected by the disease, disorder, or, any other health event).[10]

If we look at the table above and its categories, (The odds is the probability that an event will happen divided by the probability that it won't happen.). Additional labels can be added to the 2×2 table in order to better understand, and use the formula. Now, the table looks like this:

**Table 3. The 2×2 table with additional labels in the cell categories**

| | Exposed | Non-exposed | Totals |
|---|---|---|---|
| Yes | a | b | a+b |
| No | c | d | c+d |
| | a+c | b+d | a+b+c+d |

The formula for the odds ratio is:

**Odds ratio** = ad/bc = odds of exposure (cases) = ad/odds of exposure (controls) bc

In general terms, the categories represented with letters above in the table mean the following:
The letter *'a'* represents the odds of = number of cases with the exposure
The letter *'b'* represents the odds of = number of controls with the exposure
The letter *'c'* represents the number of cases without the exposure
The letter 'd' represents the number of controls without the exposure

### Example of how to calculate the Odds Ratio (or, OR)

A study hypothesized that smoking is associated with lung cancer. Initially, the investigators found 100 cases of lung cancer, and identified 100 controls free of lung cancer from the general population, similar regarding age, sex, and socioeconomic

10. CDC. (n.d.). Lesson 3: Measures of Risk. Section 5: Measures of Association In Principles of Epidemiology in Public Health Practice, Third Edition. From https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section5.html

status. At the end of the study, the entire set of numbers were released (and published), which were the following: from a total of 130 smokers, 90 of them had lung cancer.

To complete the calculation, enter first the numbers provided above, the rest of the numbers can be found by subtracting, the complete 2×2 table is shown below:

Smoking and Lung cancer

| | **Diseased (lung cancer)** | **Non-diseased** | **Total** |
|---|---|---|---|
| Exposed (smoked) | 90 | 40 | 130 |
| No exposed (no smoked) | 10 | 60 | 70 |
| Total | 100 | 100 | 200 |

**To calculate the odds ratio, use the formula provided in the content before.**

Odds ratio **(OR)** = ad/bc = odds of exposure (cases) = ad/odds of exposure (controls) bc = (90×60)/(40×10) = 5,400 /400 = 13.5

Based on the results of the OR and keeping in mind that an OR=1 implies no association. Answer the following questions: was the study hypothesis rejected, or accepted? **Is smoking a significant risk factor for lung cancer?** Answer the question based on your results of the OR.

**Relative Risk (commonly abbreviated as RR)**

Now that the odds ratio has been defined. There is need to explain, and define what the relative risk is. A comparison between the odds ratio, and the relative risk will help to understand the concept of relative risk; so, it can be said that, the odds ratio assess the risk (or, association) in general, but when there is time to start assessing the possible causation (not, only the association of a risk factor). That is the reason for the relative risk to be commonly used for the assessment or, measurement of the results of cohort studies (that are used to assess *causation*). ***The relative risk shows***

***how much more likely (or less likely) it is for people exposed to a factor to develop a disease compared to people not exposed to the factor***.[11]

**F**or the calculation of the relative risk (RR), the concept of 2×2 tables (discussed before) is used, and also, the concepts of exposure (or, exposed), and not exposed (the value in the columns of the table); and also to prepare for the calculations, the disease status or, Yes, and No. See below the layout of the 2×2 table prepared for the calculation of the relative risk:

**Table 4. The 2×2 table with additional labels in the cell categories**

| | Exposed | Non-exposed | Totals |
|---|---|---|---|
| Yes | a | b | a+b |
| No | c | d | c+d |
| | a+c | b+d | a+b+c+d |

**The formula for the relative risk is:**

**Relative Risk = (a/a+b)/(c/c+d) = the ratio** of the disease among people exposed to the factor and those not exposed to that factor.

In general terms, as we said for the odds ratio, the categories represented with letters above in the table mean the following:
The letter *'a'* represents the odds of = number of cases with the exposure
The letter *'b'* represents the odds of = number of controls with the exposure
The letter *'c'* represents the number of cases without the exposure
The letter 'd' represents the number of controls without the exposure

**Example for the calculation of the relative risk**

The same information (from the table in the odds ratio example above) – numbers can be used to calculate the relative risk formula provided before.

11. Tenny S, Hoffman MR. Relative Risk. [Updated 2021 Mar 30]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK430824/

Relative Risk **(RR)** = (a/a+b)/(c/c+d) = (90/130)/(10/70) =0.692 /0.142 = 4.8

Note: the results of the relative risk are more moderate (lower) than the results of the odds ratio. For most epidemiologists, this would be considered better because it does not exaggerate the results, making the results more credible.

Applying the information provided by the calculation of the relative risk, and keeping in mind that an RR= 1 implies no association. Answer the following questions: was the study hypothesis rejected, or accepted? **Is smoking a significant risk factor for lung cancer?** Answer the question based on your results of the RR.

*Key Takeaways*

Although the categories are the same as for the odds ratio, *the main difference is that in the calculation of the relative risk*, **the totals are used for the computation of the results.**

### Applying what we had learned so far

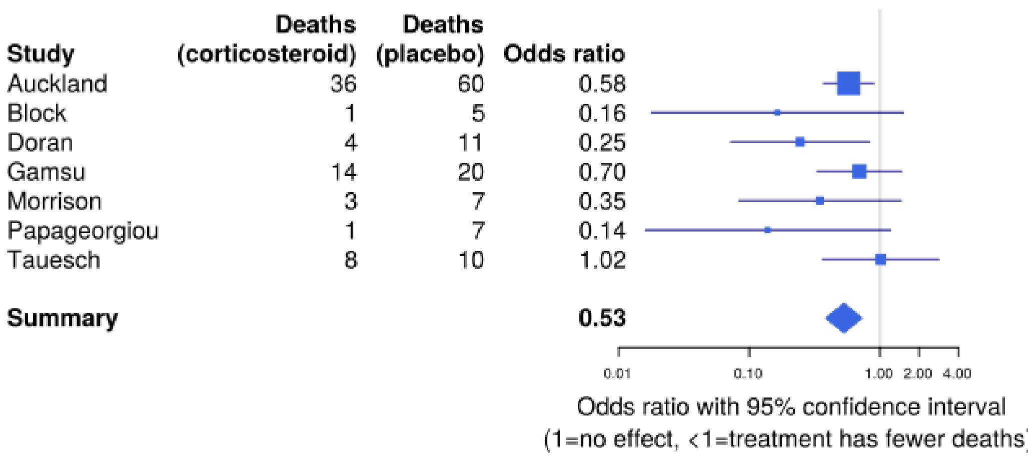
Putting together the concepts learned so far in this chapter, the meaning (for interpretation of the results) of the calculations, or, computation of both, odds ratio, and relative risk can be better understood by putting attention to the statements below:

**Interpreting the results of the odds ratio (OR) and the relative risk (RR):**

- **OR or, RR = 1.0** (or close to 1.0) means the risk of disease is similar in the exposed and unexposed group and exposure is not associated with disease.
- **OR or, RR > 1.0** means the risk of disease is greater in the exposed than the unexposed group and the exposure could be a risk factor for the disease.
- **OR or, RR < 1.0** means the risk of disease is less in the exposed group than the unexposed group and the exposure could be a protective factor.

Note: this information can be found in several sources of information available in the Internet, but there is one suggested website that discusses the topic well, and it is available at a document from Boston University, School of Public Health (see citation).[12]

Graphic example of how this information can be used and interpret especially for clinical purposes.




"Pre-term_corticosteroid_data", image by HLHB, Licensed as GNU (General Public License), Wikimedia Commons.

Note:
This is a plot chart showing information reported by several clinical trials of the use of corticosteroids to hasten lung development in preterm babies. The data used to create the plot is from a 1980s review and it does not contain up-to-date information. It is used only to illustrate the information about practical uses of the odds ratio (and in some cases, relative risk) calculations.[13]

## More applications

12. Boston University, School of Public Health. (n.d.). Measures of association. From https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_association/ep713_association_print.html
13. Roberts D, Brown J, Medley N, Dalziel SR. (2017). Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Database of Systematic Reviews 3, Art. No.: CD004454. DOI:10.1002/14651858.CD004454.pub3. Accessed 14 April 2022. From https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD004454.pub3/full

Additional notes on the application of the concepts learned in this chapter about the odds ratio and the relative risk include the concept of Attack rate, Secondary Attack rate and Case Fatality rate. Each concept will be discussed in detail as follows.

**Measures of disease outbreaks**

The concepts of odd ratio and relative risk are commonly used to investigate disease outbreaks. This type of analysis are one of the ways in which epidemiology is well-known especially for food contamination outbreaks. Also, the analysis of for example food outbreaks has tremendous financial implications, companies can lose credibility and loss of revenue if their products became contaminated. Actually some companies disappeared in some regions of the U.S. after a serious food outbreak, as it is the case of Jack in the Box fast food restaurants chain in 1993.[14]

The first of the calculations to measure the impact of disease outbreaks is the **Attack Rate.**[15] The formula for its calculation is:

> **Attack rate** = ill/(ill + well) x 100 during a time period

As simple as it look, the formula above is used as a public health measure that has been followed to control disease, separating the ill from those who are not ill or well. The separation of groups into these two categories more than useful for the analysis also help to understand why some people in the group or population did not get sick.

This apparently simple and practical formula provides results that pull together with additional information will provide the epidemiologist or investigator (of the outbreak) with meaningful information that can be put together in a chart for additional analysis, and make predictions about how to start the control of the outbreak as part of the public health team mobilization until the laboratory results

14. Food Safety News (FSN). (n.d.). Jack in the Box E. coli Outbreak – 25th Anniversary https://www.foodsafetynews.com/2017/12/jack-in-the-box-e-coli-outbreak-25th-anniversary/
15. Pettygrove, S. (n.d.). attack rate. Encyclopedia Britannica. From https://www.britannica.com/science/attack-rate

are available. The laboratory results confirm the agent that caused the disease outbreak.[16]

### Attack rate example

The Champions Elementary School held a potluck picnic for the fourth grade class. Invitations were sent to students and their parents. The day of the picnic, 135 fourth graders and 315 parents attended. The menu contained the following items: Greek salad, potato salad, dinner rolls, fruit salad, lasagna and chocolate chip cookies. It was found that of the parents, 275 ate all of the foods and 125 students ate all the foods. From this group, it was then reported that 45 students came down with staphylococcus food poisoning as did 65 parents. From the group of people who got sick 90 ate the potatoe salad, and 105 ate lasagna. Additional information: there were a group of fifty persons who had drink only and ate no food. Now, use the above information to calculate the following:

**Crude Attack rate** = ill/(ill + well) x 100 during a time period = 110/450 x 100 = 24.4% (this is called, crude because for the denominator, it counts all people invited to the picnic without excluding those who did not eat, and only drink, which are essentially not at risk, or, not susceptible).

***Key Takeaways*** *The concept of attack rate in addition, reflects or integrate the concept of relative risk, that is the reason it is said that, the relative risk is the ratio of the attack rates.*

**Attack rate (100%)** = ill/(ill + well) x 100 during a time period =x 100 = 110/400 x 100 = 27.5% (Note than in this case, the denominator is more accurate because the 50 people who did not eat and only drink were subtracted from the total of 450 who constituted the total population).

With the information above, the food specific attack rate can be also calculated, for practical reasons, only two of the food items will be used for the following calculation:

16. CDC. (n.d.). Lesson 3: Measures of Risk. Section 2: Morbidity Frequency Measures In Principles of Epidemiology in Public Health Practice, Third Edition. From https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section2.html

**Food Specific** (Potato salad) **Attack Rate** (%) = ill/(ill + well) x 100 during a time period = 90/400 x 100 = 22.5%

Food Specific (Lasagna) Attack Rate (%) = ill/(ill + well) x 100 during a time period = 105/400 x 100 = 26.25%

If the two results are compared, it can be seen that more people got sick from eating the lasagna than the potato salad. But this is not the end of the story, the food specific calculation needs to be done for the rest of the items in order to compare the results, and make a prediction of what food items caused the staphylococcus food poisoning in the mentioned picnic. The word, 'prediction' is used because the confirmation of the suspected source of the contamination needs to be confirmed by laboratory analysis samples of the food consumed during the event.

Using the concept of case discussed in another section of this textbook, the study of those ill and well can be followed by using what is known as the **Secondary Attack rate**. These are the individual who become exposed to the agent that caused the disease outbreak by being in contact directly or indirectly with the original diseased group also called, the *exposed group*. The formula that follows reflect these concepts:

**Secondary Attack rate (100%)** = # of new cases in the group minus (-) the initial case(s)/# of susceptible persons in the group minus (-) the initial cases(s) x 100

<u>Key:</u>

Initial case(s) = these are composed by the Index case(s) + co-primaries
Index case(s) = this refers to the case that first comes to the attention of the public health authorities.
Co-primaries = these are the cases related to the index case so closely in time that they are considered to belong to the same generation of cases.

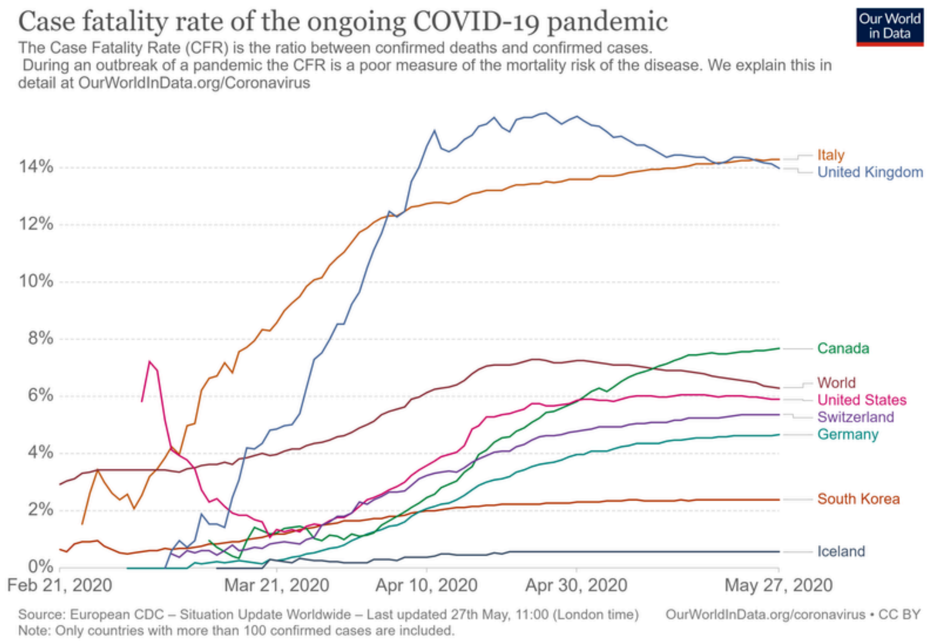
As seen above, the concept of secondary attack rate integrates the concept of Attack rate itself, and it also reflects the concepts of prevalence and incidence learned in previous sections of the book.

**Example of Secondary Attack rate**

Referring to the information of The Champions Elementary School held a potluck picnic used before in the calculation of the attack rate, the following additional information was obtained: from the people who were in the picnic, it was found that 35 took leftover potato salad home and a total of 95 families ate it the next day. Of the people in this group of families, 55 family members came down with a similar staphylococcus food poison symptoms after eating the leftover potato salad. Some of them got ill the same day, or, later in the evening or, the next day. Fifteen other family members got an upset stomach and were ill about 3 days later. With this information, calculate the secondary attack rate:

**Secondary Attack rate (100%)** = # of new cases in the group minus (-) the initial case(s)/# of susceptible persons in the group minus (-) the initial cases(s) x 100 = (110-55)/(400-55) = 55 /345 = 15.9%

As a third measure of disease outbreak, there is the **Case Fatality Rate**, which is a percentage (%). This is used to assess the capacity of an agent or any other factor to kill the affected group or, population. It is used to assess the concept of *virulence*, which has been also discussed at the beginning of the textbook. If this agent or, factor is able to kill a high number of the population, it is said that it is highly virulent. For example compared to the initial COVID 19 virus serotype, the Delta variant appears to be more virulent that the previous.




"Case Fatality Rate", Our World in Data, Licensed CC BY-SA 4.0 International, Wikimedia Commons.

The formula for case fatality rate is:

**Case Fatality Rate or CFR (%)** = # of deaths due to disease "X"/# of cases of disease "X" x 100 during a time period

At certain point in the COVID-19 pandemic, the following data were available: A certain point in the pandemic, in the U.S. there was a total of 3,542,602 confirmed cases of COVID-19, and 211,944 deaths, with this information the calculation of the Case Fatality Rate follows:

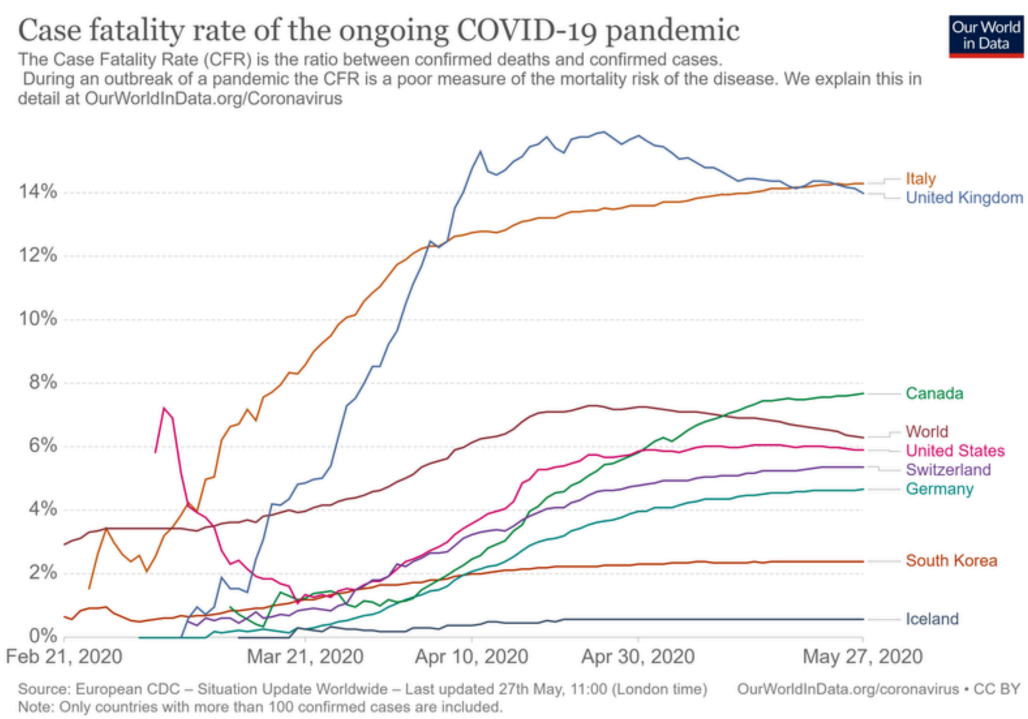
| Confirmed | Deaths | Case-Fatality | Deaths/100K pop |
|---|---|---|---|
| 3,542,602 | 211,944 | 6.0% | 651.93 |

The **Case Fatality Rate or CFR (%)** = # of deaths due to disease "X"/# of cases of disease "X" x 100 during a time period = #deaths due to COVID-19/#of cases (confirmed) of COVID-19 x 100 = 211,944/ 3,542,602 x 100 population = 5.98 (rounded 6) per 100 people. If the 100,000 pop is used as the multiplier then the result will be 5, 982.72.

It is important to note that the numbers have been changing as more cases have been confirmed, the most recent information for the U.S. is a CFR of 1.2% (CFR as March 23, 2022) = 973,213/79,803,443 x 100 = 1.21% of the population. And, the mortality per 100,000 people is 296.49 (973,213/332,403,650 x 100,000 population = 292.78 per 100,000 pop.[17]

79,803,443 973,213

As part of this information and knowing that although the data changes constantly, if it is always useful to see how calculation of the of the Case fatality rate by country in the world is presented in the publications and the news, see image below:




"Case Fatality Rate", Our World in Data, Licensed CC BY-SA 4.0 International, Wikimedia Commons.

## Application

How these concepts learned above about attack rate, secondary attack rate **are used during an infection disease outbreak?**

17. Johns Hopkins School of Medicine. (n.d.). Mortality Analyses. From https://coronavirus.jhu.edu/data/mortality

The calculations of the attack rate are used in the context of a set of procedures (also called, protocol) that make the results of these calculations meaningful. For this purpose tables are created to pull all of each individual result as a whole, and then, based on the difference, the investigator arrives to a hypothesis that will eventually lead to the finding of the possible causal agents.

To understand this process,  the CDC has published the following information also called, **protocol** for the **investigation of Infectious disease outbreaks.** These CDC's guidelines are summarized below:

**Epidemiologic Steps of an Outbreak Investigation**

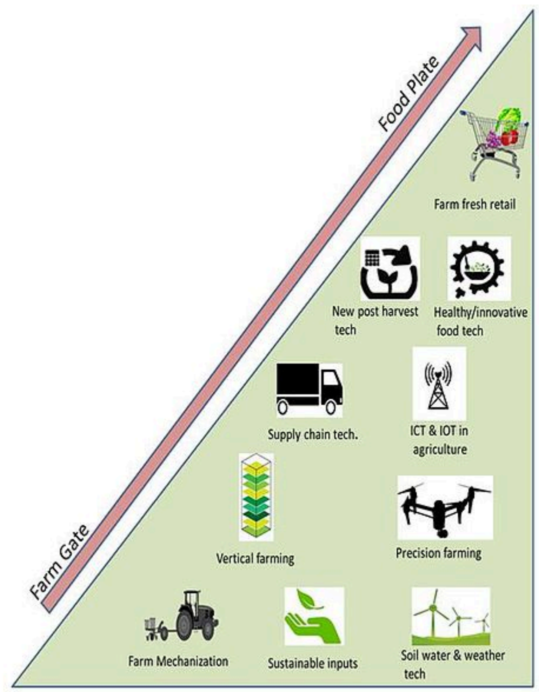
1. Prepare for field work
2. Establish the existence of an outbreak
3. Verify the diagnosis
4. Construct a working case definition
5. Find cases systematically and record information
6. Perform descriptive epidemiology
7. Develop hypotheses
8. Evaluate hypotheses epidemiologically
9. As necessary, reconsider, refine, and re-evaluate hypotheses
10. Compare and reconcile with laboratory and/or environmental studies
11. Implement control and prevention measures
12. Initiate or maintain surveillance
13. Communicate findings

*Note: This information is taken directly from CDC available at: https://www.cdc.gov/csels/dsepd/ss1978/lesson6/section2.html complete reference is found at the list of references for this chapter under the title: "Investigating an Outbreak, Steps of an Outbreak Investigation" by CDC.*

By looking at the information provided above, it is clear that the investigation of an infectious disease outbreak is a systematic process, which means, steps cannot be skipped because one step leads to another until the findings and the communication strategy is done. To illustrate the application of these principles, the following information about some infectious diseases that commonly cause illness in the community is provided.

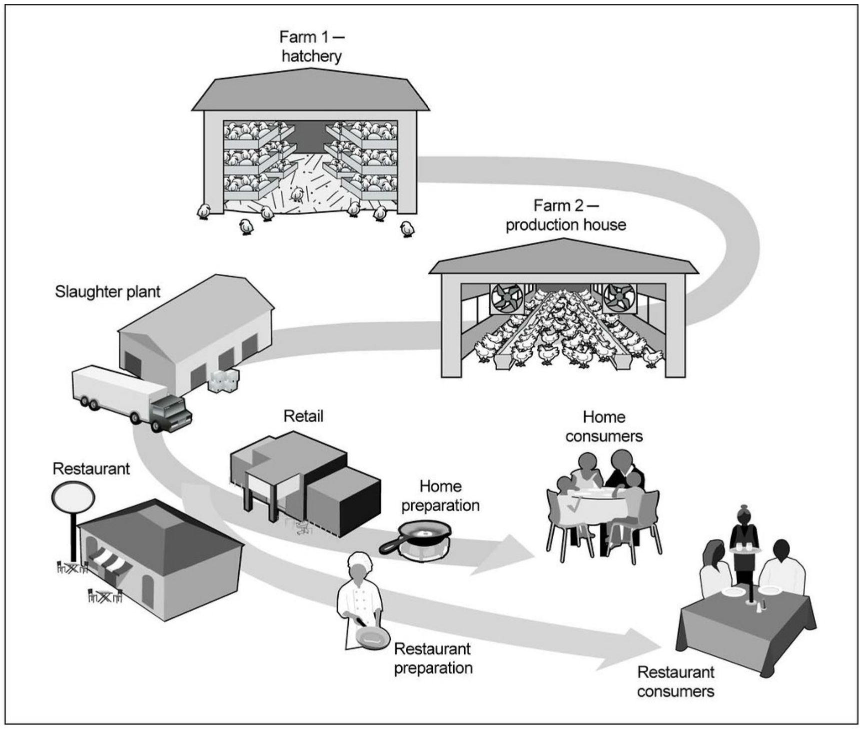
**Foodborne Illness**

This is a typical example of in which disease outbreaks are very common. It kind of make sense since before the food is served at the table, it has followed a long process from the field and the production/processing/packaging plants to the distributors and consumers. This process is commonly known as the **food production chain**. See an illustration of this process below:




"Focus Areas" (for food production safety), Licensed CC BY-SA 4.0, Wikimedia Commons.

The idea behind this approach is to put attention to the different places in which the food is has been planted (for example, fruit and vegetables), packed, and distributed even before it is at the consumer's plate. Besides the image above, the following is an specific example:




Sources: GAO and Centers for Disease Control and Prevention. | GAO-14-744

"Farm to table continuum", Public domain, Image from Flickr.

The image can be also found as part of the Food Safety: USDA Needs to Strengthen Its Approach to Protecting Human Health from Pathogens in Poultry Products Report.[18]

Several **pathogens** or, **agents**. In general, the information the investigator (clinician, epidemiologist, or, any other public health worker dedicated to this area of foodborne illness) should look for is the following:

| |
|---|
| Disease or Agent name |
| Incubation period of the disease |
| Mode of Transmission |

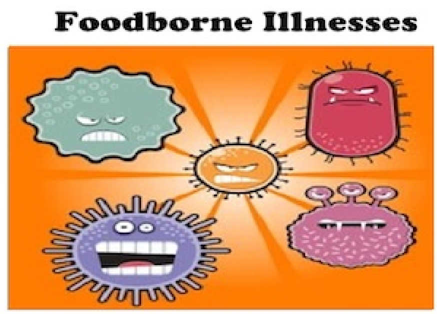
See how this information is usually found in textbooks, and the Internet. It is customary to create a table of agents, which I have re-created from several sources

18. U.S. Government Accountability Office (GAO). Food Safety: USDA Needs to Strengthen Its Approach to Protecting Human Health from Pathogens in Poultry Products. From https://www.gao.gov/products/gao-14-744

of information that are mainly of public domain. The following table contains information about the most common agents, and it does not pretend to be an exhaustive list of foodborne agents.

| Disease/Agent | Incubation Period | Mode of Transmission | Common Manifestations |
| --- | --- | --- | --- |
| Staphylococcus aureus | 2-4 hours | Contaminated food that contains staphylococcus enterotoxin | Gastrointestinal syndrome, majority of cases with vomiting |
| Cholera/Vibrio cholerae | 2-3 days | Contaminated water that contains infected feces or vomitus, also contaminated food | Profuse watery diarrhea (painless) |
| Campylobacter enteritis/ Campylobacter jejuni | 2-5 days | Abdominal pain, diarrhea pain, malaise, fever | Undercooked chicken or pork, contaminated food and water, raw milk |
| Clostridium perfringens food poisoning | 10-12 hours | Heavily contaminated food (for example, meats and gravies that have been inadequately heated or stores especially at temperatures that allows bacteria multiplication | Diarrhea |
| Clostridium botulinum/Classic botulism | 12-36 hours | Contaminate food containing toxins (for example, home canned foods) | Classic syndrome compatible with botulism that include neurological symptoms besides the diarrhea. |
| Salmonellosis/Various species of Salmonella (for example, S. typhimurium and S. enteritidis) | 12-36 hours | Contaminated food that contains Salmonella organisms (for example, undercooked chicken, eggs, meat, raw milk) | Gastrointestinal syndrome |

**Table of Common Foodborne Agents**, created by the author of this textbook, Giovanni Antunez, 




"Foodborne Illnesses", Image from SafetyNews

Common foodborne agents are staphylococcus aureus, clostridium botulinum and Escherichia coli (commonly known just as E. coli). But, depending on the source of the report and the year of publication, common bacteria foodborne agents include, nontyphoidal *Salmonella* spp., *Clostridium perfringens*, and *Campylobacter* spp. Leading causes of hospitalization were nontyphoidal *Salmonella* spp., norovirus (not include in the table because it is a virus and not a bacteria) and *Campylobacter* spp. [19]

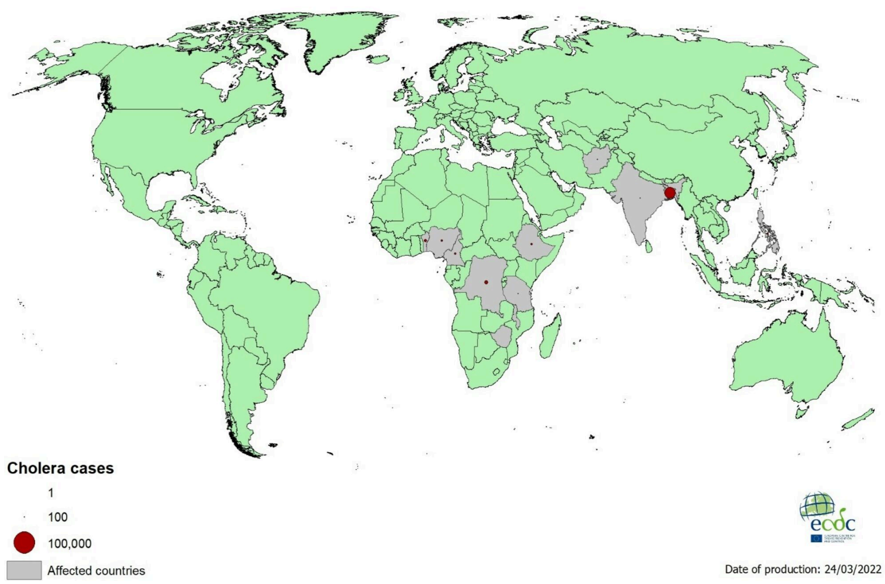
## Waterborne diseases

Some of the waterborne diseases are also foodborne as it is the case of Cholera, which is transmitted mainly by contaminated drinking water, and also by contaminated food. Common waterborne diseases besides cholera include, giardiasis, amebiasis that are intestinal parasites, and legionellosis (caused by bacteria) and schistosomiasis, which is a blood parasite transmitted by water contaminated with human waste).

19. Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M., Roy, S. L....Griffin, P. M. (2011). Foodborne Illness Acquired in the United States—Major Pathogens. Emerging Infectious Diseases, 17(1), 7-15. https://doi.org/10.3201/eid1701.p11101

### Cholera

Once epidemic, the last epidemic of cholera in the new world, which is also known as the Americas is, the epidemic of cholera in Peru, Mexico and Central America in the 1990s.[20] A disease of the past century came back as a re-emerging disease in those mentioned countries. and also in recent years in Africa, one of the most dramatic is the cholera epidemic in Liberia and Sierra Leone in Africa just after the Ebola epidemic. And, in recent reports, cholera continuous to be a problem for several countries, below is a map of the affected countries:

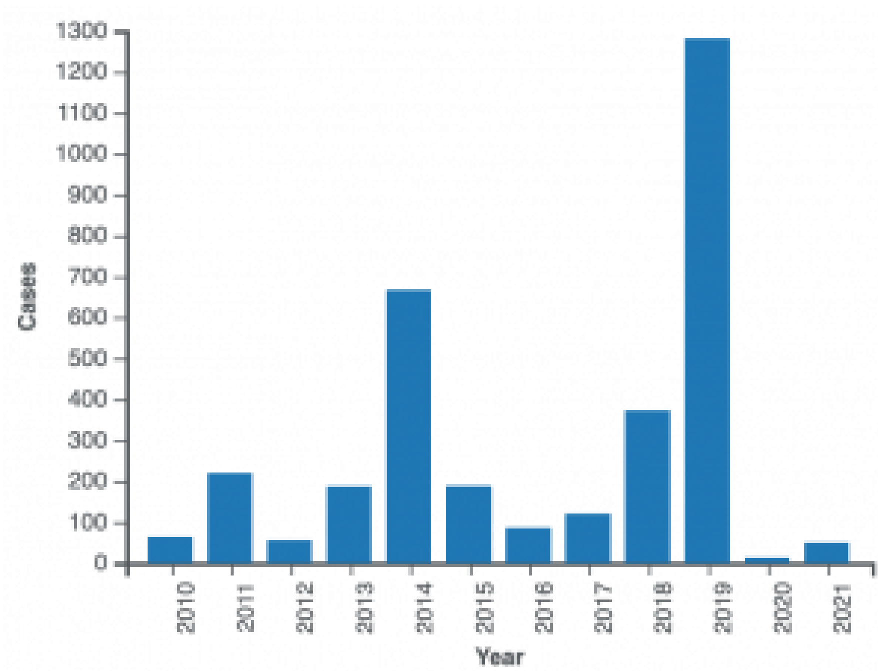



[“Geographical distribution of new cholera cases reported worldwide, January – March 2022”,](#) Image from [The European agency of Prevention and Control](#).

20. Olsvik O. Koleraepidemien i Latin-Amerika [The cholera epidemic in Latin America]. Tidsskr Nor Laegeforen. 1992 May 30;112(14):1843-6. Norwegian. PMID: 1631846.

### Vaccine Preventable Diseases

Before the current COVID-19 pandemic, vaccine preventable diseases had been the cause of major disease outbreaks in the world not to say epidemics. The following are common vaccine preventable diseases: diphtheria, pertussis, tetanus, Haemophilus influenzae different serotypes, hepatitis A and B, measles, rubella, mumps, paralytic poliomyelitis. Classically, the majority of these diseases had affected the world's children population, and because of the major emphasis in child's vaccination, these diseases although not completely eradicated yet, they remain mostly under control. But cases of disease outbreaks especially of measles have been reported in recent years, as it is the case.




"Reported measles cases in the United States by year, as of February 20, 2022", image from CDC.[21]

### Sexually Transmitted Diseases

Other diseases that commonly generate disease outbreaks, includes HIV/AIDS which constitutes the most significant example of a sexually transmitted disease that reached epidemic proportions since the 1980s, and although no major attention has been brought to this serious public health problems, HIV/AIDS is still a serious

21. CDC. (2022). Measles Cases and Outbreaks. From https://www.cdc.gov/measles/cases-outbreaks.html

problem in the United States and the world. The following is a map of the estimated number of people living with the disease in the world:

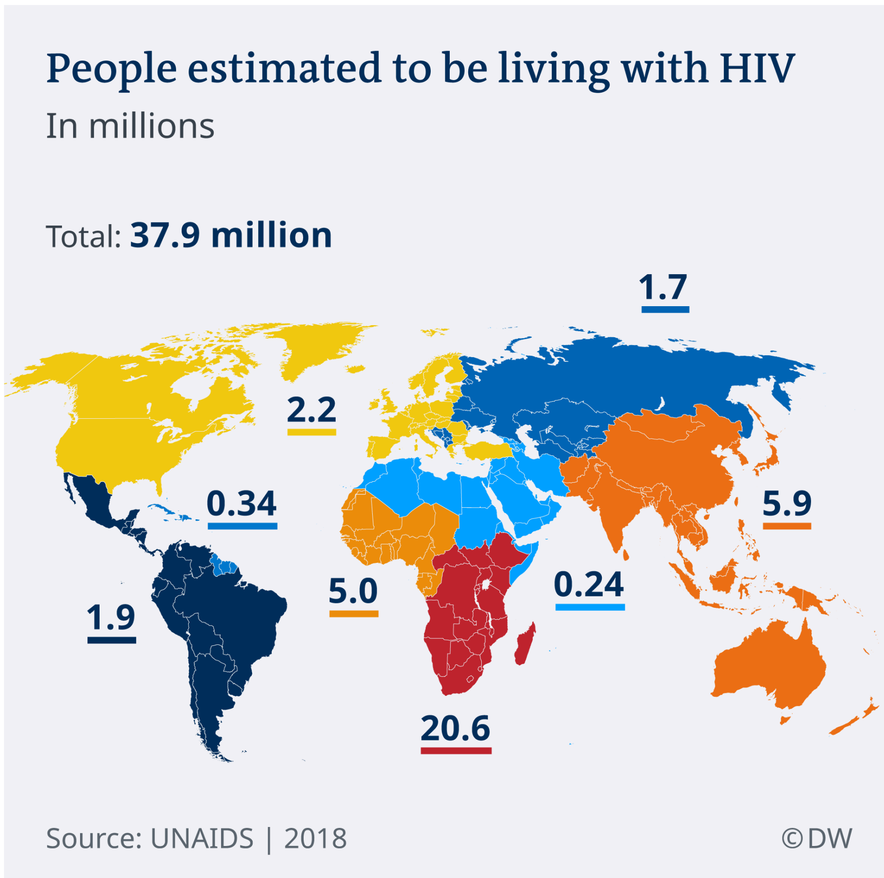



Image from [DW (Deutsche Welle) News]

The HIV/AIDS pandemic shows that there is need to remind us that although COVID-19 is the most pressing current health situation in the world, there is also other infectious and non infectious diseases that are waiting for attention of the world health care system. Otherwise, the world will enter in a continuous health emergency that will be difficult to control if no preparations are taken before re-emergent and new emerging diseases appear.

**Summary.** This chapter had covered the concepts of odds and relative risk and how they are used in epidemiology, and public health. Examples of both forms of assessment have been provided, and the second part of the chapter covered mainly applications and calculations that are used commonly for the investigation of disease outbreaks. These latter calculations are mainly applications of the concepts of prevalence and incidence, which is the topic of the previous chapter.

# Glossary

## Epidemiology Glossary[1]

**This glossary has been taken directly from the CDC web site:** https://www.nwcphp.org/docs/risk/epi_glossary.pdf

In most cases, when you alphabetically search for a term/definition, this book takes you to the original source or, the CDC website. To return to this book, click on the back arrow in your browser.

- A | B | C | D | E | F | G | H | I | L | M | N | O | P | R | S | T | U | V | Y | Z

- **AGE-ADJUSTED MORTALITY RATE.** A mortality rate statistically modified to eliminate the effect of different age distributions in the different populations.
- **AGENT.** A factor, such as a microorganism, chemical substance, or form of radiation, whose presence, excessive presence, or (in deficiency diseases) relative absence is essential for the occurrence of a disease.
- **AGE-SPECIFIC MORTALITY RATE.** A mortality rate limited to a particular age group. The numerator is the number of deaths in that age group; the denominator is the number of persons in that age group in the population.
- **ANALYTIC EPIDEMIOLOGY.** The aspect of epidemiology concerned with the search for health-related causes and effects. Uses comparison groups, which provide baseline data, to quantify the association between exposures and outcomes, and test hypotheses about causal relationships.

1. Centers for Disease Control and Prevention (CDC). (2014). Epidemiology Glossary, from https://www.cdc.gov/reproductivehealth/data_stats/glossary.html

- **ANALYTIC STUDY.** A comparative study intended to identify and quantify associations, test hypotheses, and identify causes. Two common types are cohort study and case-control study.
- **APPLIED EPIDEMIOLOGY.** The application or practice of epidemiology to address public health issues.
- **ASSOCIATION.** Statistical relationship between two or more events, characteristics, or other variables.
- **ATTACK RATE.** A variant of an incident rate, applied to a narrowly defined population observed for a limited period of time, such as during an epidemic.
- **ATTRIBUTABLE PROPORTION.** A measure of the public health impact of a causative factor; proportion of a disease in a group that is exposed to a particular factor which can be attributed to their exposure to that factor.