

# Methodology in epidemiology

**Epidemiology** has as its research tool the so-called **epidemiological work method**. It is a research method together with a clinical, biological, experimental method that has three basic procedures: descriptive, analytical and experimental. Depending on what we want to find out with the given procedure, whether the cause or occurrence of the disease, or to describe it, we distinguish different methods.

## Descriptive studies

**Descriptive studies** are observational studies, they describe the *distribution of the disease in the population* (according to characteristics person, place and time) and compare their occurrence in different geographical territories, in of different races, nations, ethnicities and social groups, in different time periods.

- They are a source of hypotheses, **indicating** to a possible causal (causal) **relationship** between various factors and the development of disease.
- These are observational studies that collect, sort and evaluate data on disease and mortality for a given disease. Of all the characteristics in this discipline, age is the variable that is most important and must be taken into account as much as possible.
- The goal of descriptive characteristics in medicine is to **show the relationship and connection** between living conditions, disease and death and the consequences caused by the above variable characteristics.
- The causal relationship between mortality, environmental and lifestyle factors, provides epidemiologists with the opportunity to develop **prevention and control programs** to suppress a given disease.
- When causality is understood, there will be places where it is possible to **intervene** so that the given phenomenon in the population decreases and the health level of the population increases.
- Mostly in them we monitor incidence, prevalence, mortality from a given disease in population groups in relation to various characteristics of a person, place and time.
- Mausner and Bahn propose to include in descriptive epidemiology as basic variables, in addition to age, also place and time as basic concepts used to describe events and activities that surround us or can cause disease outbreaks.

## Basic epidemiological characteristics

In an epidemiological study, we talk about three basic characteristics:

1. characteristics of the person,
2. location characteristics,
3. characteristic of time.

### Characteristic of a person: the question WHO?

These are characteristic characteristics of the monitored persons:

- **age**,
  - we can express the relationship between morbidity and age:
    - using **simple age curve** (captures all inhabitants of different age groups in 1 moment in time),
    - using the **cohort curve** (shows indicators for groups of people born at the same time and followed repeatedly in subsequent calendar periods - cohorts are important if morbidity changes over time),
  - some diseases are typical for certain age groups:
    - **newborns** - congenital defects, perinatal infection,
    - **children** - childhood rash diseases, rotavirus infections,
    - **young adults** - STDs, drug addictions, Multiple Sclerosis, Crohn's Disease,
    - **elderly people** - CVD, cancer, nosocomial infections, degenerative diseases,
- **gender**,
  - examples of different occurrence:
    - **men** - CVD, lung cancer, atherosclerosis, ulcer disease, injuries,...
    - **women** - cholecystitis, diabetes mellitus, thyrotoxicosis, obesity, psychoneuroses,...
- **employment**,
- **ethnic group**,
  - blacks - sickle cell disease,
  - Caucasians - multiple sclerosis,
  - Eskimos - lip cancer,
- **education**,
- **socioeconomic status**,
  - in the lower social strata there is a higher incidence of STD, ca cervix, due to increased promiscuity and reduced awareness of these groups of people,
- **marital status**
  - loneliness, or single status, widowhood are increasingly associated with the rate of psychological disorders, such as depression and with the occurrence of suicides,
- **nutritional status**,
- **personal history, family history**.

## Place characteristics: the question WHERE?

- **natural conditions**
  - geographic factor,
    - altitude – high a.s.l. = frequent polyglobulia,
    - climate – air temperature and humidity (decisive e.g. for carriers of infections),
    - chemical and physical factors,
- **social conditions**
  - lifestyle, presence of pollutants, population density,
    - *For example: cirrhosis of the liver in France, together with the French paradox - high consumption of wine,*
- **size of the monitored area**
  - a nursery, or a city, or a landscape
- **characteristic of environment**
  - isolation or overpopulation of the environment,
- **dominant site factor**
  - is marked when **all** ethnic groups living in a given place **get the same disease** and the ethnically mixed population of another area has a low incidence and [[prevalence|prevalence] ] of the given indicator,
    - immigrants to the given territory fall ill,
    - after the given incubation period, emigrants from the given territory become ill,
    - animals will also fall ill in the given territory (in the case of an infection also transmissible to animals),
- **cartogram** = map with marked place of occurrence of the observed disease.

## Characteristics of time: the question WHEN?

- **secular development** (trends)
  - long-term trends of change – they take place slowly and continuously in one direction,
- **changes in progress**
  - curve of permanent rise or permanent fall, or stationary curves,
- **periodic changes** (daily, weekly, seasonal, annual, long-term,...),
- **irregular changes**
  - explosive epidemic – highly infectious agent, toxic chemicals.

## Data scales

**Descriptive statistics** is a discipline that quantitatively describes the main properties of a collection of data, e.g. measurement results. Descriptive (also descriptive) statistics try to summarize essential information about the given data with a few numbers and pictures. Descriptive statistics is used by e.g. epidemiology.

Descriptive statistics deals with characters (data) of different nature. Since the character of the character can affect the way in which the statistical file can be described, the data is divided into several **data groups**, which we can also call **data scales**. So, for example, for data that has a nominal nature, it does not make sense to describe it using an average or median, only modus can be used as a position measure.

### Nominal data

- These are data that are **only descriptive**, they cannot be sorted and if any numbers are assigned to them, it is only to mark them in a certain way.
- These are essentially all epidemiological characteristics (place, time, gender...).
- If we were to take, for example, ethnicity, we have white, black, Hispanic, Asian, Indian...

### Ordinal data

- Data that already has a **certain order** (from the English *order* = order).
- For further analysis, it can sometimes be useful to combine the data into categories, e.g. if we code the respondents' education with codes starting from zero (no education) gradually according to the level of education achieved, we can, for example, group bachelor's and master's education for the purposes of analysis.
- An example is the generally known *probability scale* from 1 to 5:

if, for example, we were to ask in a health questionnaire how many people could be vaccinated if they opened a vaccination station in the neighborhood, we could sort their answers as follows:

1. he would not go to get vaccinated;
2. would consider vaccination;
3. maybe he should go get vaccinated;
4. he would probably go get vaccinated;
5. he would certainly go and get vaccinated;

### Interval data

- It provides **more information** than nominal and ordinal scales, specifically, unlike ordinal data, it makes sense to evaluate data not only in the sense of "larger-smaller", but also in the sense of "how much one value

is greater than the other".

- Does not have a fixed zero value - the choice of zero is somewhat arbitrary, although there may be good reasons for it.
- Example: IQ scale - The average is 100, this Intelligence Quotient value is the majority of the population. Above-average intelligence has values above 100. Just because someone has an IQ of 0 doesn't mean they have no intelligence, but that the range of this scale is set so that it can't accommodate such a low level of intelligence.

## Ratio data

- The scale on which the ratio data is distributed is often referred to as the ratio scale.
- For this scale, the **zero point is already fixed** and expresses the complete absence of the given value.
- Unlike the interval scale, it also makes sense to talk about how many times one value is larger/smaller than the other.
- Example: weight, mortality indicator, often the number of points from the test.

## Formulation of hypotheses

### Definition

- A **hypothesis is an assumption** whose **validity** we must verify before proceeding with further scientific work.
- Provisional theory is based on descriptive research/statistics, clinical observations, analytical studies, laboratory research, theoretical modeling.

### Hypothesis formation methods

#### Differential method

- If the frequency of occurrence of a disease in two comparative populations (sets) is **significantly different** and if it is possible to identify a factor that is present in one set and not in the other, then we can label this factor as **cause of disease**.
- *Example: **higher incidence of ca of the cervix in married women - and vice versa** lower<sup>1</sup> incidence in nuns.*

#### Match Method

- If we can find a **common factor** in a number of different populations that are characterized by the occurrence of a disease, then this factor could be the **cause of the disease**.
- *Example: Semmelweis came to the conclusion that the cause of teenage fever was some kind of "dead poison" found 'on the hands of the attending obstetric staff.*

#### Method of analogy

- If there are *certain similarities* in the distribution of the observed disease with another disease for which we have more complete information, it would be possible to judge that both diseases have *some common causes*. We use a **deductive way of thinking**.

#### Companion Difference Method

- The change in the frequency of the disease occurs in parallel with the *change in intensity* of the given factor.
- *Example: incidence of **lung cancer** increases with the number of **cigarettes smoked**.*

## Links

- ws:Metodologie v epidemiologii

## Related articles

- Epidemiology

## Reference

## References

- 
-