Epidemiology

19

Elisabeth Presterl, Magda Diab-El Schahawi,
Luigi Segagni Lusignani, Helga Paula, and Jacqui S. Reilly

Epidemiology is the study of the occurrence of diseases, their course and their distribution in a population. The term itself is derived from the Greek epi = "upon" and demos = "population".

- "Epidemiology is the science whose object is the distribution and propagation of diseases in human populations" [1].

The origin of epidemiology lies in loimology (study of pestilential diseases and plagues). Among them was understood the doctrine of detection, control and prevention of communicable diseases. The British physician John Snow recognized in 1854, on the occasion of a cholera epidemic in London, that this disease was not spread by fumes (miasmas), as was widely accepted at the time. Almost all sufferers got their water from a certain pump on Broad Street. This well had been contaminated by sewage with microorganisms in an attempt to flush the open, foul-smelling sewers into the Thames. A cholera epidemic with 14,000 dead was the result. John Snow was able to prove that the deaths centred around a water pump on Broad Street. After he put the pump out of operation, it came to a standstill of the epidemic. However, his theory was not recognized during his lifetime by the then scientists and doctors and confirmed only some years after his death [2].

E. Presterl (✉) · M. Diab-El Schahawi (✉) · L. S. Lusignani · H. Paula
Department of Hygiene and Infection Control, Medical University of Vienna, Vienna, Austria
e-mail: elisabeth.presterl@meduniwien.ac.at; magda.diab-elschahawi@meduniwien.ac.at;
luigi.segagnilusignani@meduniwien.ac.at; helga.paula@meduniwien.ac.at

J. S. Reilly
Glasgow Caledonian University, Glasgow, UK
e-mail: jacquelinereilly@nhs.net

With the increasing importance of chronic noninfectious diseases, the work of epidemiologists has also changed. Modern epidemiology also includes noninfectious diseases, their causes and risk factors (e.g. obesity, cardiovascular diseases, cancer).

The general aim of epidemiology is to describe/study the distribution of the incidence of diseases in a population by identifying etiological factors, providing data for planning and utilizing the knowledge gained for occupational or population health purposes.

Data Sources

Epidemiology draws its data from different sources:

- *Primary data*: Data collected specifically for examination purposes (e.g. health survey of the German Cardiovascular Prevention Study, Bundesgesundheitssurvey 1998).
- *Secondary data*: Obtained from primary data through modelling or processing steps. The majority of existing health-related data is secondary data (i.e. data on health insurance benefits, such as incapacity to work).
- *Causes of death*: Official statistics based on the death certificate. Only the main cause of death is included in the statistics. The reliability (reliability) of this data depends on care and knowledge of the underlying cause of death.
- *Types of illness*: Social security benefits, in particular incapacity for work, hospital diagnostic statistics and pension benefits due to occupational or occupational disability (only suitable as epidemiological morbidity measures).
- *Collection of notifiable diseases*: Epidemic law; due to various reporting obligations (illness, suspicion), limited data are available, which is also related to the problem of reporting discipline.
- *Disease register*: Registration of all persons suffering from a particular illness or who have died (regionally available for a few diseases in Austria). Examples: cancer registries.

19.1 Methods and Measures of Epidemiology

Descriptive Epidemiology: describes the disease pattern of a population as well as those characteristics that may be related to this distribution pattern. Here, no or only very fundamental statistics are performed; it is the description of the data in the foreground.

Question: Who? Where? What? When?

Analytic Epidemiology: results in hypotheses of disease development, which are being investigated in the population. For this purpose, statistical methods and mathematical models are used to obtain an interpretation from the data obtained. Here one can distinguish between qualitative analysis ("Are there indications of interrelationships between variables?") and quantitative analysis ("How strong are these relationships?"). This allows statements on causality of/between risk factors.

Question: Why? How?

19.2 Measurements of Epidemiology

In epidemiology, numbers are used either as *absolute numbers* (are not very meaningful because the reference value is missing) or as *relative numbers* (rates, the absolute number is related to a constant population, i.e. mortality, lethality, etc.).

The most important measures in epidemiology are:

- Prevalence
- Incidence
- Mortality
- Letality
- Morbidity

Prevalence is the number of all "existing" cases of a particular disease "at a given time". It is a measure of the incidence of disease *in a defined population* at a given time.

$$\text{Prevalence} = \frac{\text{Nr. of all existing cases of a particular disease at a specific point in time}}{\text{Nr. of all persons of the total population at risk at a specific point in time}}$$

Influencing factors:

- Number of new cases
- Disease duration
- Case definition (depending on diagnostics methods)
- Migration (inflow and outflow)
- Causes of disease

Variants:

- Point prevalence
- Period prevalence

Example: how many students of the "X" Faculty have a cold?
Population at risk: students of the "X" Faculty
Given time: today
Population: 40
People with cold: 10

$$\text{Prevalence} = \frac{10}{40} = \textbf{25\%} \rightarrow \text{Prevalence of cold of the "X" Faculty students today is 25\%.}$$

The calculated measure is an example of descriptive epidemiology: the population and the desired characteristics are merely described, but no statements can be made, e.g. about the onset of the disease or the morbidity rate.

This type of prevalence is also called point prevalence because it is a single point in time. In contrast to:

$$\text{Period prevalence} = \frac{\text{Nr. of all existing cases of a particular disease at a specific period of time}}{\text{Nr. of all persons of the total population at risk at a specific period of time}}$$

Period prevalence measures the number of existing cases in a defined population over a predetermined period.

Incidence is the number of new cases of illness related to a specific number of persons (population) in a given period of time. It measures the new cases within a given period of time in a defined group of individuals who were free of the disease at the beginning of the observation period. The initially disease-free group is also called "population under risk". Within incidence, the cumulative incidence risk is to be distinguished from the incidence density.

$$\text{Incidence} = \frac{\text{Nr. of new cases in a period of time}}{\text{Population exposed to the risk in this period to time}}$$

Example: how many students of the "X" Faculty will have a cold in the next week?

Population at risk: students of the "X" Faculty
Given time: 7 days
Population: 40
People with new cold: 4

$$\text{Incidence} = \frac{4}{40} = 10\% \rightarrow \textbf{Incidence of cold of the "X" Faculty students is 10\% per week.}$$

The incidence rate represents the magnitude of morbidity within a population, as it takes into account the period during which disease-free individuals were "risk-exposed" to a disease.

Morbidity is the frequency with which a disease appears in a population. Morbidity can be an incidence measure (number of persons in a population who become ill (incidence)) or a period prevalence (number of persons who are ill at a given time from the morbidity rate; the disease probability can only be estimated).

Mortality is a measure of the frequency of occurrence of death in a defined population (e.g. 10,000, 100,000 persons) during a specified interval

$$\text{Mortality}\left(\text{crude death}\right) = \frac{\text{Nr. all deaths}\left(\text{year}\right)}{1000 \text{ individuals in the population}}$$

$$\text{Mortality}\big(\text{cause-specific}\big) = \frac{\text{Nr.deaths attributed to a specific cause}\big(\text{year}\big)}{1000\,\text{individuals in the population}}$$

Letality (or *case fatality rate*) is the number of deaths over the number of sick with a specific disease. It is the proportion of cases in a designated population of a particular disease, which die in a specified period of time. Lethality represents a measure of risk. It is a better measure of clinical significance of the disease than mortality, and it is most often used for diseases with discrete, limited time courses, such as outbreaks of acute infections.

19.2.1 Standardization

In order to make a comparison between epidemiological rates, they have first to be standardized. Standardization is intended to "balance" distorting structural differences in the population and may, for instance, be applied for age, gender or other characteristics. Standardization by age is particularly prevalent, as information is usually available and age is important in most health problems.

Age standardizations based on a standard population are often used in cancer registries to compare morbidity or mortality rates. If different age structures are present in populations of different regions or in the population of an area over time, their mortality or morbidity rates are only limitedly comparable. For interregional or intertemporal comparisons, age standardization is necessary. Here, the reference population is the age structure of a reference population, the so-called standard population assumed. The age-specific mortality or morbidity rates of the reference population are weighted according to the age structure of the standard population. After age standardization, data from different years or regions can be compared with each other without causing any distortion due to different age structures. When interpreting age-standardized morbidity or mortality rates, it should be noted that they do not represent real (in the sense of empirically observable) information. Rather, they describe what the mortality or morbidity rates in the considered population would be if the reference population corresponded to the standard population, i.e. was abstracted from age structure-related effects.

19.3   Epidemiologic Studies

19.3.1 Cross-Sectional Studies

Indicators: the frequency of diseases (prevalence) and the simultaneous occurrence of risk factors or exposure parameters are recorded in one population at a time, so that—in contrast to the following case-control study—generally valid statements on the association between the diseases and the exposure, also expressed as relative risks, are possible. Relative risk is the factor by which the probability of disease increases (or decreases) under exposure. The statement of a cross-sectional study is descriptive.

Period: the duration of the study is relatively short; a time sequence (exposure → illness) cannot be proven. Especially with smaller populations, possible selection effects (i.e. avoidance behaviour of particularly sensitive persons in relation to hazardous substances or also the withdrawal from the region particularly affected) must be taken into account when interpreting the results.

Population: in contrast to the ecological study, individuals are directly measured so that the relationship between the diseases and the exposure to other factors can be controlled.

19.3.2 Case-Control Studies

Indicators/Population: in this type of study, a group of previously identified patients is retrospectively examined for the presence of the risk factors and the exposure compared to a non-affected control group. From the relative frequency of exposure in patients or non-sufferers, the so-called odds ratio can be taken as an estimate of the relative risk of the disease under the exposure. This type of study allows for evidence of causality.

Period: defined period (prospective or retrospective).

Especially in cancer epidemiology, this type of study is very often used, because the studies provide relatively fast results compared to the long latency of the diseases; furthermore case-control studies should be conducted with huge observational numbers in order to ensure statistically proved statements. However, a serious disadvantage of these studies lies in the information about the true exposure status of cases and the control persons in the past, which is often distorted by various factors.

19.3.3 Cohort Studies

While in the case-control study the direction of the disease is exposure, in the cohort study, it is always exposure to the disease.

Period: In a (prospective) cohort study, a population that differs in the exposure status of individuals is monitored over a defined period of time, for example, to detect the onset of disease as a function of exposure status.

Indicators: Incidence (the incidence of new diseases) and mortality from a particular disease (for subgroups of the population, e.g. for "heavily exposed"). From the possibilities of statement, this type of study, in which parallel exposure status and disease probability are considered, is the most versatile, provided that the observation periods are sufficiently large enough to be able to show effects at all.

Special case retrospective cohort study: data on the course of exposure development and the target diseases for the study population are already available and will be analysed retrospectively. However, this type of study is not flexible in the study design; it is not possible to easily incorporate new exposure parameters, disturbance

variables or target diseases into the observation phase (since the data were collected in the past), so this study type also requires suitable exposure and registry data. The statements of a cohort study are mostly incidence and causal statements.

Population: A defined population (e.g. groups, families, villages, regions) is sub-divided into exposed and nonexposed, and both groups are compared for the proportion of already "ill" and "not ill".

While in the case-control study the direction of the disease is to the exposure, in the cohort study, it is always from the exposure to the disease.

19.3.4 Intervention Studies

The starting point of an intervention study is a cohort study, whereby in a subpopulation the exposure status has changed due to an external intervention in the observation period, so that the effect of this intervention on the development of the target diseases can be observed (e.g. comparative studies between real drug and placebo). This type of study comes closest to an experimental study (control and control of key exposure parameters).

19.4 Analytical Measures

19.4.1 Relative Risk

Relative risk (RR) can be calculated from cohort studies showing the disease incidences of exposed and unexposed persons. For this purpose, a contingency table is created with the absolute values of the respective group.

	Disease		
Exposure	Yes	No	
Yes	a	b	a + b
No	c	d	c + d
	a + c	b + d	

The risk of the exposed cased is R(EX) = a/a + b (cumulative incidence in the exposure group).

The risk of the not exposed cases is R(NEX) = c/c + d.

If both values are divided, one obtains a measure indicating the probability of an event occurring in an exposed group to the probability of the event occurring in a comparison, nonexposed group.

The relative risk is:

$$RR = R(EX)/R(NEX) = (a/a+b)/(c/c+d)$$

Exercise: do people with hypertension have an increased risk of coronary heart disease? This involves comparing a control group with an experimental group and investigating whether there is any effect to the study group. This effect is called "relative risk".

Exposure	Coronary heart disease		
	Yes	No	
Yes	43	1475	1518
No	69	11,635	11,635
	112	13,153	

19.4.2 Odds Ratio

Odds ratio (OR) is defined as the ratio of the probabilities (odds, chances) of an event occurring in one group to the probabilities of it occurring in another group. Odd ratios show the correlations between exposure and disease in case-control studies.

Odds = "Chances"; Odds Ratio = "relative Chances"

OR is very similar to the relative risk, but does not include any incidence, but rather prevalence differences between exposed and nonexposed persons. It is used to figure out if a particular exposure is a risk factor for a particular outcome and to compare the various risk factors for that outcome.

Exposure	Disease		
	Yes	No	
Yes	a	b	a + b
No	c	d	c + d
	a + c	b + d	

$$OR = a^* d / c^* b$$

(Measure of the strength of a difference between groups → sets a relation between odds of exposed and not exposed groups)

- Relation between ill and not ill under exposure
- Relation between ill and not ill with no exposure

Odds ratios can therefore be interpreted as a measure of interrelation:

- OR = 1 means that there is no difference in odds.
- OR > 1 means that odds of the exposed group are higher.
- OR < 1 means that odds of the exposed group are lower.

19.5   Outbreak Management and Basic Epidemiology

By definition, an outbreak is the cumulative occurrence of infectious diseases where an epidemic link is likely or suspected. Outbreaks can occur in the form of:

- *Pandemic*: limited in time, spatially unlimited (e.g. influenza during the winter months)
- *Endemic*: unlimited in time, spatially limited (e.g. norovirus within a nursery)
- *Epidemic*: temporally and spatially limited (e.g. cholera within several neighbourhoods)

Sporadic individual cases of illness (for instance, individual norovirus sufferers in some hospital wards) are not an outbreak.

The goal of an outbreak investigation is to quickly identify the cause(s) and prevent further transmissions. Outbreaks are usually recognized by meticulous healthcare workers. Eighty percent of all reported outbreaks are no outbreaks, but 80% of all outbreaks are only detected by scrupulous reporting and tracking!

The basis of outbreak management is to determine whether, in the event of an outbreak, the cases have a common cause or source. The "tools" of outbreak management are descriptive and analytical epidemiology, microbiological examinations and, last but not least, the use of common sense. An outbreak investigation must be structured and conscientious. The steps of such an outbreak investigation are generally structured as follows as soon as the notification of an unusual incidence of disease cases or MDRO comes in:

1. Prompt general control measures.
2. Confirm outbreak.
3. Form outbreak team.
4. Site visit.
5. Case definition (secure diagnosis).
6. Determine cases (line list).
7. Collect data (time/place/person).
8. Hypothesis formulation.
9. Analytical study on hypothesis.
10. Targeted control and prevention measures (detection of infection source).
11. Create a report.
12. Surveillance.

The creation of an epidemic curve is an instrument for the visualization of the temporal-/spatial-/population-related relationships of the outbreak. Such a curve is a visual aid to document and track the course of an outbreak. The curve can be easily created by pen and paper, and no complicated technical tools are necessary.

Special outbreak programs are available, which are a useful tool, especially in pandemics, and the creation of a Microsoft Excel document to track the progression of increased prevalence of infection is often used (see Example).

|				x			x	10 cases
				x				
				x				
			x	x				
			x	x				
	1	2	3	4	5	6		
	2017							

Example: episodes of *Staphylococcus epidermidis* isolated in blood cultures from Ward X (January to June 2017)

On the basis of the distribution pattern of the epidemic curve, the nature of the outbreak can sometimes already be recognized as in Fig. 19.1.

19.5.1 Case Definition

A case in an outbreak is defined by:

- *Place* (ward, OP, ICU, etc.)
- *Time* (day, month, admission, microorganisms, etc.)
- *Person* (gender, age, disease, etc.)
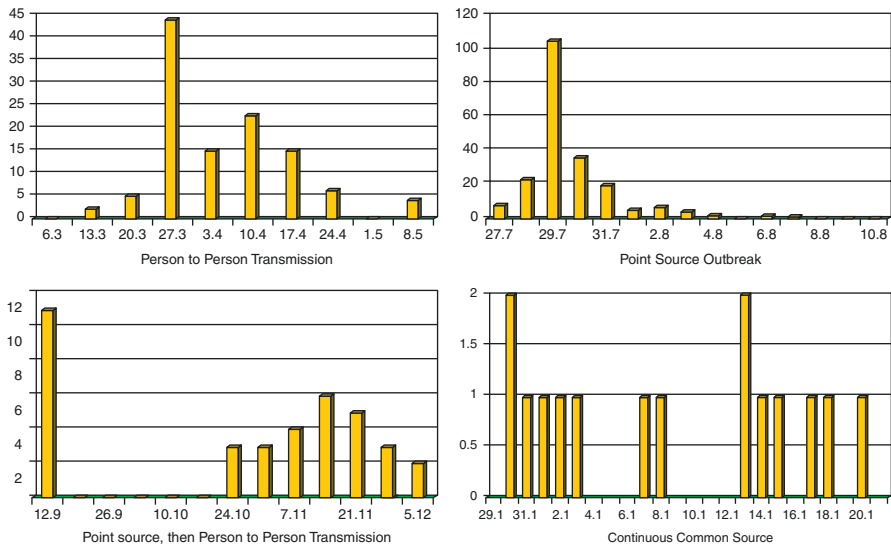
Fig. 19.1 Possible distribution pattern of the epidemic curve

19.5.1.1 Time-Person-Place

At the end of an outbreak investigation, the event is analysed retrospectively by the outbreak management team. Gained information should be used to implement control and prevention measures or to complement existing measures. An *outbreak report* should be prepared in which the case and the procedure for examination, diagnostics, evaluation, etc. are described in detail. The final evaluation should include deficit analysis and definition of future prevention strategies.

Appropriate questions are as follows:

- Was a timely detection of the outbreak ensured?
- Did the breakout management team and communication chains work efficiently?
- Were the immediate targeted measures correctly taken?
- Have any further illnesses occurred despite the measures taken?
- Was an efficient cause clarification ensured by hygienic, microbiological and epidemiological investigations?
- Was a causal clarification of the sources of infection and chains of infection ensured?
- Which prevention strategies have been proven effective?
- Which prevention strategies had to be modified or newly established?

References

1. Frentzel-Beyme R. Einführung in die Epidemiologie. Darmstadt: Wissenschaftliche Buchgesellschaft; 1985.
2. Snow J. On the mode of communication of cholera. 2nd ed. London: John Churchill; 1855.