

ABC OF DEFINITIONS AND TERMS

J.P. Vaughan, R.H. Morrow

The definitions given in this chapter are valid as they are used in this publication but different definitions may be used in other contexts. This chapter is largely based on *A dictionary of epidemiology*, edited by J.M. Last for the International Epidemiological Association and published by Oxford University Press, 1983.

A

Access. The proportion of a defined population that has a particular facility within reasonable reach, which may be measured by distance, time, costs or social and cultural factors.

Accuracy. The degree to which a measured value represents the true value of the variable that is being measured. See Repeatability and Validity.

Agent. A factor whose presence or deficiency is essential for the occurrence of a disease, e.g. micro-organisms, chemical substances, vitamins and essential amino acids.

Age-sex pyramid. See Population pyramid.

Age-specific rate. A rate for a specified age group, with the numerator and denominator for the same age group.

$$\text{e.g. 1-4 years} = \frac{\text{Number of deaths among 1-4 year old children in area in one year}}{\text{Average total population aged 1-4 years}} \times 100$$

In ← same ← area ← in ← same ← year

Airborne infection. A disease caused by an infectious agent capable of being transmitted by particles or droplets suspended in the air, e.g. measles and pertussis.

Arbovirus. A group of diverse animal viruses that are transmitted to humans by blood-feeding arthropod vectors, such as mosquitoes, ticks, sandflies and midges. The term is an abbreviation of "arthropod-borne virus".

Association. Statistical dependence between two or more variables, which are said to be associated if they occur together more frequently than would be expected by chance. Statistical tests enable the degree of association to be calculated.

Attack rate. This rate usually refers to the incidence of new cases during an epidemic. The secondary attack rate is based on the number of new cases among contacts of a primary case that occur within the accepted incubation period of the disease. The denominator is the total number of exposed contacts during the same period of time.

Average. See Mean, arithmetic

B

Bias. Any effect during the collection or interpretation of information that leads to a systematic error in one direction, e.g. errors resulting from weighing scales under-recording a child's true weight, or observer bias in the interpretation of replies to questions in a questionnaire.

Birth rate. A summary crude rate based on the number of live births in a known population over a given period of time.

$$\text{Birth rate} = \frac{\text{Number of births in area during on year}}{\text{Average total population in area during the same year}} \times 100$$

Birth weight. Infant's weight recorded at the time of birth. Low-birth-weight babies weigh less than 2500 grams and the percentage of such babies is commonly used as a general measure of health status.

C

Carrier. A person or animal that has a specific infectious agent in the absence of clinical disease and that serves as a potential source for the further transmission of the infection.

Case. A person who is identified as having a particular characteristic such as a disease, behaviour or condition. An epidemiological definition of a case is not necessarily the same as the clinical definition. Cases may be divided into possible, probable and definite, depending on how well specific criteria are satisfied.

Case-control study. An analytical epidemiological study that compares cases of a particular condition with suitable control subjects, who do

not have the condition, looking at the frequency of associated factors in the two groups. Sometimes also called a retrospective study. Often used to test hypotheses about etiology, e.g. the link between lung cancer and cigarette smoking.

Case-fatality rate. The percentage of persons contracting a disease who die from it. This rate is most commonly used for communicable diseases.

$$\text{Case-fatality rate} = \frac{\text{Number of deaths from a disease in a given period}}{\text{Number of cases of disease diagnosed in the same period of time}} \times 100$$

Catchment area. The geographical area from which the people attending a particular health facility come.

Census. The enumeration of an entire population, usually with details being recorded on residence, age, sex; occupation, ethnic group, marital status, birth history and relationship to head of household. A *de facto* census only counts the people who are actually present during the enumeration, whereas a *de jure* census records all people by their normal place of residence at the time of enumeration.

Chemoprophylaxis. The administration of drugs to prevent infection from occurring or to prevent the infection from progressing into disease.

Class. A group of observations made on a variable, considered together for the convenience of analysis, e.g. haemoglobin values may be classed by intervals of 1g/dl.

Clustering. The grouping of a series of cases in relation to time or spacial area or both. The space time clustering of cases in an epidemic commonly indicates a point-source outbreak due to an infectious agent or toxic chemical.

Cluster sampling. A sampling method in which each unit selected is composed of a group of persons rather than an individual, e.g. villages and households.

Cohort. A well defined group of people who have had a common experience or exposure, who are then followed up for the incidence of new disease or events, as in a cohort or prospective study. A group of people born during a particular period or year is called a birth cohort.

Communicable period. The time during which an infectious agent may be transferred from an infected person to another susceptible person, or from an animal to man or vice versa.

Confounding. A situation in which the effects of two variables are difficult to separate from each other, e.g. level of family income and availability of food as causes of malnutrition.

Contact. Exposure to a source of an infection. Transmission due to direct contact may occur when skin or mucous membranes touch, as in body contact, kissing and sexual intercourse.

Contagious. Transmitted by contact or close proximity.

Control. Disease control programmes aim to lower the incidence of new cases, or reduce the proportion of severe cases through treatment, to an acceptably low level, so that the disease is no longer considered a major public health hazard.

Control group. Comparison group of people who do not have a particular disease or condition or who have not been exposed to the disease, intervention, procedure or other variable that is being studied. Neighbourhood controls, which are commonly used for convenience, are people who live in the same neighbourhood. See also Case-control study.

Correlation. A measure of association that indicates the degree to which two or more sets of observations fit a linear or straight-line relationship. Correlation may be positive, when both variables increase together, or negative, when one increases as the other decreases.

Coverage. A measure, usually expressed as a percentage, of people or households who have actually received a particular service compared to all those who need it, e.g. percentage of households with a reasonably safe water supply, percentage of infants immunized with three doses of DPT vaccine.

$$\text{Birth rate} = \frac{\text{Number of births in area during on year}}{\text{Average total population in area during the same year}} \times 100$$

Cross-sectional survey. A survey or study that examines people in a defined population at one point in time. Cross-sectional surveys usually supply prevalence data but repeated surveys can be used to give an estimate of incidence.

D

Data processing. Conversion of raw data into a form suitable for analysis with computers and statistical programmes.

Death rate. The proportion of a population who die from any cause during a specified period of time. The rate can be made specific for a particular cause, or group of causes, of death. The rate can also be calculated for each sex and for any age group, thus providing disease, sex and age-specific rates.

$$\text{Crude death rate} = \frac{\text{Number of all deaths during one year}}{\text{Average total population in during same year}} \times 100\%$$

Demand for health care. Willingness and/or ability to seek and use services. Sometimes further divided into expressed demand or actual use and potential demand or need.

Demography. The study of populations, with reference to such factors as size, age structure, density, fertility, mortality, growth and social and economic variables.

Denominator. The lower portion of a fraction. In the calculation of rate, this represents the total population at risk.

Disease, subclinical. The condition in which a disease is only detectable by special tests and there are no apparent symptoms and signs.

E

Endemic. The constant presence of a disease or infectious agent in a given population or geographical area. Also used to refer to a disease with a constant incidence of new cases in the area.

Epidemic. The occurrence in a community or region of cases of an illness or other similar event clearly in excess of what is normally expected. The characteristics of the illness, the area and the season all have to be taken into account. To judge whether there is an excess or not requires knowledge about the previous incidence of the event in the same area.

Epidemiology. The study of the distribution and determinants of health and disease in populations

and its application to the prevention and control of health problems and diseases.

Eradication. The extermination of an infectious agent, thus halting transmission of infection e.g. smallpox has been eradicated throughout the world and malaria has been eradicated from certain areas.

Evaluation. A process that attempts to determine as systematically and objectively as possible the relevance, effectiveness and impact of activities in the light of their objectives. Evaluation is often carried out separately for inputs, processes, outcomes and impact.

Expectation of life. The average number of years an individual is expected to live if current mortality trends continue. Life expectancy at birth is the average number of years a newborn baby can be expected to live under existing conditions. As many deaths in developing countries occur during infancy and childhood, the average life expectancy in these countries is much lower than in developed countries.

F

False negative. A false result in a screening test, leading to the classification of a person, who is actually positive, as negative or normal.

False positive. A positive test result in a subject who is actually negative, i.e. a healthy person is wrongly said to have a particular disease or attribute.

Fertility rate. See General fertility rate.

Fetal death rate. Also called stillbirth rate. The number of fetal deaths in one year expressed as a proportion of all births (live plus stillbirths) in the same year.

$$\text{Fetal death rate} = \frac{\text{No. of fetal deaths in one year}}{\text{No. of fetal deaths plus live births in same year}} \times 1000$$

G

General fertility rate. Similar to the crude birth rate, except that the denominator is restricted to women of childbearing age, i.e. 15-44 years.

$$\text{General fertility rate} = \frac{\text{No. of live births in area in one year}}{\text{Average female population aged 15-44 years for same area and year}} \times 1000$$

Growth rate of populations. Also known as the natural rate of population increase. In the absence of the effects of migration, it is calculated as the crude birth rate minus the crude death rate.

H

Health. A state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity.

Health indicator. A measure that reflects, or indicates, the state of health of persons in a defined population, e.g. the infant mortality rate.

Health information system. A combination of health statistics from various sources, used to derive information about health status, health care, provision and use of services, and impact on health.

Herd immunity. The resistance of a group or community to invasion and spread of an infectious agent, due to the resistance to infection in a high proportion of individual members of the group. The herd immunity results from the lowered probability of the disease agent being transmitted from an infected person to a susceptible one when a high proportion of individuals are not susceptible.

Holoendemic. Describes a disease that is virtually universal in the population, with symptoms in childhood, leading to a state of equilibrium and a lower incidence of symptoms in adults, e.g. malaria in some communities, especially in Africa.

Host. A person or animal that is infected under natural conditions. A number of microorganisms and parasites may have several different hosts.

Household interview survey. The collection of information from a representative sample of households by trained interviewers. It is usually a cross-sectional survey to collect information about individual members and on common features, e.g. water supply.

Hyperendemic. A disease that is constantly present at a high incidence (or prevalence) and that affects all age groups.

I

Incidence. The number of new cases or events or attendances occurring in a defined population within a given of time, commonly one year.

Incidence rate. A measure of the rate at which new cases or events occur in a defined community.

$$\text{Incidence rate} = \frac{\text{No. of new cases or events diagnosed in population in one year}}{\text{Average total population at risk in same area in one year}} \times 100\%$$

Incubation period. The time interval between invasion of a susceptible host by an infectious agent and the appearance of the first symptom or sign of the disease.

Infant mortality rate. A measure of the rate at which deaths occur in children less than one year old.

$$\text{Infant mortality rate} = \frac{\text{No. of deaths in children less than 1 year old in one year}}{\text{No of live births in same year}} \times 1000$$

M

Maternal mortality rate. A measure of a woman's risk of dying from causes associated with pregnancy. A maternal death is the death of a woman while pregnant or within 42 days of the termination of pregnancy, irrespective of the duration and the site of pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes. Some countries have extended the period of 42 days to up to one year.

Maternal deaths are subdivided into (a) direct obstetric deaths and (b) indirect obstetric deaths resulting from pre-existing disease or disease that developed during pregnancy and which was not due to direct obstetric causes, but which was aggravated by physiological effects of pregnancy. The death of pregnant woman from an incidental cause (e.g. motor car accident) is not classified as a maternal death.

$$\text{Maternal mortality rate} = \frac{\text{No. of maternal deaths in given area during one year}}{\text{No of live births in population in the same area during the same year}} \times 1000$$

Mean, arithmetic. This is also commonly called the average. It is calculated by adding together all the

individual values in a group of measurements and dividing by the number of values in the group.

Measurement scale. The complete range of possible values for a measurement. Scales can be divided into five main types:

Dichotomous - two mutually exclusive groups, such as positive and negative.

nominal - qualitative categories, such as for religions.

ordinal - ordered qualitative categories, such as social classes I to V.

interval - scale with equal distances for each interval but not particular starting or zero point, such as date of birth.

ratio - interval scale with a zero starting point, such as weight, blood pressure; income.

Median. The central value in a range of measurements that divides the set into two equal parts.

Mode. The most frequently occurring value in a set of observations.

Monitoring. The continuous measurement and observation of the performance of a service or programme to see that it is proceeding according to the proposed plans and objectives. If monitoring reveals that there are problems, management decisions will have to be taken to alter or improve the service or programme so that it comes back on track.

Morbidity. Any departure from a state of wellbeing. Morbidity can be expressed in terms of people who are ill and/or as episodes of illness.

N

Neonatal mortality rate. The number of deaths in infants under 28 days of age in a given period, usually one year, per 1000 live births in the same period.

Non-respondents. Members of a study sample or population who do not take part, respond or participate, for whatever reason, in the study. Respondents may differ from non-respondents and a high non-response rate may be an important source of bias.

Notifiable disease. A disease that, by statutory requirements, must be reported to the public health authority.

Numerator. The upper portion of a fraction. In calculating a rate, all people included in the numerator should have derived from the denominator. However, this is not true for the numerator in a ratio.

O

Observational study. Study, survey or investigation that is made by observing subjects and where no interventions, or at least no additional ones, are implemented at the same time.

Observer error. Variation or error in measurements due to failure of the observer to measure or identify the phenomenon accurately. Variation can be due to such faults as the observer missing an observation, poor technique, incorrect reading or recording, and misinterpretation of answers to questions. Observer error is particularly important if it is non-random and biased.

Output. The immediate results that come from health care or programme activities expressed as units, of service, such as number of outpatient visits or persons immunized.

P

P or probability value. The letter *P* followed by <, the symbol for less than, and a number (usually 0.05, 0.01 or 0.001) is a statement of the probability that the association or observation could have occurred by chance. The number 0.05 means the observation would be to occur by chance 1 in 20 times; similarly, 0.01 means 1 in 100. An association is commonly accepted as statistically significant if *P* is < 0.05.

Pandemic. An epidemic occurring over a very wide area.

Pathogenesis. The mechanism by which an etiological agent produces disease.

Perinatal mortality rate: The officially accepted definition is as follows:

$$\text{Perinatal mortality rate} = \frac{\text{Late fetal deaths (28 weeks or more gestation)} + \text{Plus first-week postnatal deaths}}{\text{Fetal deaths plus total live births}} \times 1000$$

In the same population over same period

However, the definition accepted in many countries that do not have good vital statistical records leaves fetal deaths out of the denominator. Perinatal mortality is a useful indicator of the quality of antenatal and obstetric care and is usually given as a rate per 1000 births per year.

Population. The total number of inhabitants of a given area or country. In sampling, the population may refer to the units from which the sample is drawn, not necessarily the total population of people. The term population is also commonly used to refer to particular subgroups, such as priority or high-risk groups.

Population pyramid. A graphical representation of the age and sex composition of a population. A pyramid with a broad base, sloping sides and narrow apex is typical of many developing countries. This shape is due to high fertility and high mortality at younger ages.

Postneonatal mortality rate. The number of infant deaths between 28 days and one year of age in a given year per 1000 live births in that year. In developing countries this rate largely reflects deaths due to infectious diseases and malnutrition.

Predictive value. The probability that a person with a positive (or negative) result in a screening or diagnostic test is in fact a true positive (or true negative). These are called the positive and negative predictive values of the test. The predictive value depends on the sensitivity and specificity of the test and on the prevalence of the condition being screened. *See* Validity.

Prevalence. The number of cases or events or conditions in a given population at a particular point in time.

Prevalence rate. The total number of cases or events or conditions at a particular point in time divided by the total population at risk at the same point in time. Prevalence rates are most commonly used for diseases or events that have a long average duration.

Prevalence study or survey. *See* Cross-sectional study.

Prevention. Measures aimed at promoting and maintaining health, by such interventions as improving nutritional status, immunization, suitable water supplies and excreta disposal (primary pre-

vention). Secondary prevention comprises measures aimed at ensuring the early detection of diseases and infections, whereas tertiary prevention is concerned with reducing symptomatic illness and disability.

R

Random. Describes a happening or event due to chance and not determined by other factors.

Randomization. The separation or allocation of individuals to two or more groups at random. Randomization should form two or more groups with variables randomly allocated between the groups.

Randomized controlled trial. An experiment using people randomly allocated to treatment or intervention groups and a control group. The results are assessed by looking for any significant difference between these groups. Such trials are the most rigorous and scientific way of testing the effectiveness of new interventions.

Random sample. A sample derived by random selection of sample units. Each individual unit, such as village, household or person, should have an equal chance of being included in the sample.

Relative risk. The ratio of the risk of death or disease in an exposed population to the risk in the unexposed population.

Repeatability. The ability of a test to produce results that are identical or closely similar each time it is conducted. Precision is another term that is often used. *See* Accuracy and Validity.

Representative sample. A sample that resembles the original population or reference population in every way. To ensure this, all chosen samples should be compared with the original population, particularly for important variables such as age and sex.

Reservoir of infection. The natural habitat of an infectious agent, which may be a person, animal, arthropod, plant, soil, etc. It is where the agent normally lives and multiplies.

Response rate. The number of interviews or examinations completed divided by the total due to have been carried out, expressed as a percentage. A high non-response rate can be an important source of bias.

Retrospective study. See Case-control study.

Risk. The probability that an event will occur, e.g. that an individual will become ill or die within a stated period of time or age. The term is usually used with reference to unfavourable events.

Risk factor. The term is used in at least two different ways: (1) an attribute, variable or exposure that is associated with an increased probability of a specified event, such as the occurrence of a disease. Such preceding factors are not necessarily causal (also called risk markers); (2) an attribute, variable or exposure that actually increases the occurrence of a specified event, and is therefore believed to be causal (also described as a determinant).

S

Sample. A selected subset of a population. A sample may be random or non-random and it may be representative or non-representative. In an *epsem* (equal probability of selection method) sample all the population units have an equal chance of being selected. A simple random sample is an *epsem* sample.

Screening. This is the presumptive identification of unrecognized disease or behaviour by using tests, examinations, questionnaires and other procedures. Screening sorts people into positives and negatives or normals. People who are positive will probably require further investigation. It is important to examine the results for the proportion of false positives and false negatives. See also Sensitivity and Specificity.

Sensitivity. The proportion of true positives correctly identified by a screening test. See Predictive value and Specificity.

Seroepidemiology. The use of serological investigations, particularly antibody levels, to detect infections and transmission patterns.

Socioeconomic status. A descriptive classification of a person's position in society, using such criteria as income, educational level, occupation and dwelling place. Attitudes towards health and health status are often closely linked to socioeconomic status.

A classification similar to the one used by the Registrar-General of the United Kingdom is as follows:

<i>Social group</i>	<i>Occupation</i>
I	professional
II	intermediate
III N	non-manual skilled
III M	manual skilled
IV	partly skilled
V	unskilled

This classification may be applicable in an industrialized society but is less useful in many developing countries.

Specificity. The proportion of true negatives correctly identified by a screening test. See also Predictive Value and Sensitivity.

Sporadic. A disease or event that occurs infrequently and irregularly. A term usually applied to certain communicable diseases.

Spot map. A map showing the geographical distribution people with a particular characteristic, commonly used in the investigation and control of an epidemic.

Standard deviation. A measure of the dispersion or variation of a set of quantitative observations or measurements on either side of the mean or average.

Standardization. Application of statistical techniques to standardize two or more populations for differences that may exist between them, particularly in the age-sex structure, to enable valid comparisons to be made.

Statistical significance. See *P* or probability value.

Stillbirth rate. See Fetal death rate.

T

Total fertility rate. An estimate of the total number of children a thousand women would bear if they went on having children at the present age-specific fertility rates. It provides an answer to the question: How many children does a woman have on average during her lifetime?

Transmission of infection. The spread of an infectious agent, either through the environment or from person to person. The main mechanisms of transmission are: direct contact, placental, fomite borne, vector-borne and air-borne.

Trend. A long-term general movement or change in frequency, usually either upwards or downwards. a downward trend in a disease or unhealthy behaviour means that it is becoming less frequent.

U

Under-reporting. Failure to identify or count all cases or events, leading to a numerator that is smaller than the true one. This leads to estimates of frequency that are lower than the true value.

V

Validity. The degree to which a measurement actually measures or detects what it is supposed to measure. This concept is particularly important in

screening procedures. See Accuracy and Repeatability.

Variable. Any characteristic or attribute that can be measured.

Virulence. The degree of pathogenicity, or ability to produce disease, of an infectious agent.

Vital statistics. Systematically tabulated information about births, marriages, divorces and deaths, based on registration of these vital events.

Z

Zoonosis. An infectious or communicable disease that can be transmitted from vertebrate animals to human beings.