

# GLOSSARY

## A

**active immunity** see immunity, active.

**active surveillance** see surveillance, active.

**age-adjusted mortality rate** see mortality rate, age-adjusted.

**agent** a factor (e.g., a microorganism or chemical substance) or form of energy whose presence, excessive presence, or in the case of deficiency diseases, relative absence is essential for the occurrence of a disease or other adverse health outcome.

**age-specific mortality rate** see mortality rate, age-specific.

**alternative hypothesis** see hypothesis, alternative.

**analytic epidemiology** see epidemiology, analytic.

**analytic study** see study, analytic.

**antibody** any of a variety of proteins in the blood that are produced in response to an antigen as an immune response.

**antigen** any substance (e.g., a toxin or the surface of a microorganism or transplanted organ) recognized as foreign by the human body and that stimulates the production of antibodies.

**applied epidemiology** see epidemiology, applied.

**arbovirus** any of a group of viruses that are transmitted between hosts by mosquitoes, ticks, and other arthropods.

**arithmetic mean** see mean, arithmetic.

**arithmetic-scale line graph** see line graph, arithmetic-scale.

**arthropod** an organism that has jointed appendages and segmented external skeleton (e.g., flies, mosquitoes, ticks, or mites).

**association** the statistical relation between two or more events, characteristics, or other variables.

**asymmetrical** a type of distribution where the shape to the right and left of the central location is not the same. Often referred to as a skewed distribution; the mean, median, and mode of an asymmetrical distribution are not the same.

**asymptomatic** without symptoms.

**attack rate** a form of incidence that measures the proportion of persons in a population who experience an acute health event during a limited period (e.g., during an outbreak), calculated as the number of new cases of a health problem during an outbreak divided by the size of the population at the beginning of the period, usually expressed as a percentage or per 1,000 or 100,000 population (see also **incidence proportion**).

**attack rate, secondary** a measure of the frequency of new cases of a disease among the contacts of known patients.

**attributable proportion** see **proportion, attributable**.

**attributable risk percent** see **proportion, attributable**.

**attribute** a risk factor that is an intrinsic characteristic of the individual person, animal, plant, or other type of organism under study (e.g., genetic susceptibility, age, sex, breed, weight).

**axis** one of the dimensions of a graph in a rectangular graph, the x-axis is the horizontal axis, and the y-axis is the vertical axis.

## B

**bar chart** a visual display in which each category of a variable is represented by a bar or column bar charts are used to illustrate variations in size among categories.

**bar chart, 100% component** a stacked bar chart in which all bars or columns are the same length, and the measured axis represents 0%–100%.

**bar chart, deviation** a bar chart displaying either positive or negative differences from a baseline.

**bar chart, grouped** a bar chart displaying quantities of two variables, represented by adjoining bars or columns (i.e., a group) of categories of one variable, separated by space between groups.

**bar chart, stacked** a bar chart displaying quantities of two variables, represented by subdivided bars or columns (the subdivisions representing the categories of one variable) separated by space between bars or columns.

**bias** a systematic deviation of results or inferences from the truth or processes leading to such systematic deviation; any systematic tendency in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth. In epidemiology, does not imply intentional deviation.

**bias, information** systematic difference in the collection of data regarding the participants in a study (e.g., about exposures in a case-control study, or about health outcomes in a cohort study) that leads to an incorrect result (e.g., risk ratio or odds ratio) or inference.

**bias, selection** systematic difference in the enrollment of participants in a study that leads to an incorrect result (e.g., risk ratio or odds ratio) or inference.

**bimodal** having two data peaks.

**biologic transmission** see **transmission, biologic**.

**birth cohort** see **cohort, birth**.

**birth rate, crude** the number of live births during a specified period divided by the mid-period population, usually expressed per 1,000 population.

**box plot** a visual display that summarizes data by using a "box and whiskers" format to indicate the minimum and maximum values (ends of the whiskers), interquartile range (length of the box), and median (line through the box).

## C

**carrier** a person or animal that harbors the infectious agent for a disease and can transmit it to others, but does not demonstrate signs of the disease. A carrier can be asymptomatic (never indicate signs of the disease) or can display signs of the disease only during the incubation period, convalescence, or postconvalescence. The period of being a carrier can be short (a transient carrier) or long (a chronic carrier).

**case** an instance of a particular disease, injury, or other health conditions that meets selected criteria (see also **case definition**). Using the term to describe the person rather than the health condition is discouraged (see also **case-patient**).

**case-control study** see **study, case-control**.

**case definition** a set of uniformly applied criteria for determining whether a person should be identified as having a particular disease, injury, or other health condition. In epidemiology, particularly for an outbreak investigation, a case definition specifies clinical criteria and details of time, place, and person.

**case-fatality rate** (also called **case-fatality ratio**) the proportion of persons with a particular condition (e.g., patients) who die from that condition. The denominator is the number of persons with the condition; the numerator is the number of cause-specific deaths among those persons.

**case, index** the first case or instance of a patient coming to the attention of health authorities.

**case-patient** in a case-control study, a person who has the disease, injury, or other health condition that meets the case definition (see also **case**).

**case, source** the case or instance of a patient responsible for transmitting infection to others; the instance of a patient who gives rise to an outbreak or epidemic.

**cause, component** a factor that contributes to a sufficient cause (see also **cause, sufficient**).

**cause of disease** a factor (e.g., characteristic, behavior, or event) that directly influences the occurrence of a disease. Reducing such a factor among a population should reduce occurrence of the disease.

**cause, necessary** a factor that must be present for a disease or other health problem to occur.

**cause-specific mortality rate** see **mortality rate, cause-specific**.

**cause, sufficient** a factor or collection of factors whose presence is always followed by the occurrence of a particular health problem.

**census** the enumeration of an entire population, usually including details on residence, age, sex, occupation, racial/ethnic group, marital status, birth history, and relationship to the head of household.

**central location** (also called **central tendency**) a statistical measurement to quantify the middle or the center of a distribution. Of the multiple ways to define central tendency, the most common are the mean, median, and mode.

**chain of infection** the progression of an infectious agent that leaves its reservoir or host through a portal of exit, is conveyed by a mode of transmission, and then enters through an appropriate portal of entry to infect a susceptible host.

**"chartjunk"** unnecessary or confusing visual elements in charts, illustrations, or graphs. The term was first used by Edward Tufte in his book, *The Visual Display of Quantitative Information* (1983).

**class interval** the span of values of a continuous variable that are grouped into a single category (see also **class**), usually to create a frequency distribution for that variable.

**class limits** the values at the upper and lower ends of a class interval.

**clinical criteria** the medical features (e.g., symptoms, medical examination findings, and laboratory results) that are used in a case definition.

**clinical disease** a disease that has been manifested by its symptoms and features.

**clinical trial** see **trial, clinical**.

**cluster** an aggregation of cases of a disease, injury, or other health condition (particularly cancer and birth defects) in a circumscribed area during a particular period without regard to whether the number of cases is more than expected (often the expected number is not known).

**cohort** a well-defined group of persons who have had a common experience or exposure and are then followed up, as in a cohort study or prospective study, to determine the incidence of new diseases or health events.

**cohort, birth** a group of persons born during a particular period or year.

**cohort study** see **study, cohort**.

**common-source outbreak** see **outbreak, common-source**.

**community immunity** see **immunity, herd**.

**community trial** see **trial, community**.

**comparison group** a group in an analytic study (e.g., a cohort or case-control study) with whom the primary group of interest (exposed group in a cohort study or case-patients in a case-control study) is compared. The comparison group provides an estimate of the background or expected incidence of disease (in a cohort study) or exposure (in a case-control study).

**confidence interval** a range of values for a measure (e.g., rate or odds ratio) constructed so that the range has a specified probability (often, but not necessarily, 95%) of including the true value of the measure.

**confidence limits** the end points (i.e., the minimum and maximum values) of a confidence.

**confounding** the distortion of the association between an exposure and a health outcome by a third variable that is related to both.

**contact** exposure to a source of an infection; a person who has been exposed.

**contact, direct** exposure or transmission of an agent from a source to a susceptible host through touching (e.g., from a human host by kissing, sexual intercourse, or skin-to-skin contact) or from touching an infected animal or contaminated soil or vegetation.

**contagious** capable of being transmitted from one person to another by contact or close proximity.

**contingency table** a two-variable table of cross-tabulated data.

**continuous variable** see **variable, continuous**.

**control** in a case-control study, a member of the group of persons without the health problem under study (see also **comparison group** and **study, case-control**).

**crude** when referring to a rate, an overall or summary rate for a population, without adjustment.

**crude birth rate** see **birth rate, crude**.

**crude death rate** see **mortality rate, crude**.

**crude mortality rate** see **mortality rate, crude**.

**cumulative frequency** in a frequency distribution, the number or proportion of observations with a particular value and any smaller value.

**cumulative frequency curve** a plot of the cumulative frequency rather than the actual frequency for each class interval of a variable. This type of graph is useful for identifying medians and quartiles and other percentiles.

## D

**death-to-case ratio** the number of deaths attributed to a particular disease, injury, or other health condition during a specified period, divided by the number of new cases of that disease, injury, or condition identified during the same period.

**decision analysis** application of quantitative methods to decision-making.

**decision tree** a branching chart that represents the logical sequence or pathway of a clinical or public health decision.

**demographic information** personal characteristics of a person or group (e.g., age, sex, race/ethnicity, residence, and occupation) demographic information is used in descriptive epidemiology to characterize patients or populations.

**dendrogram** see **phylogenetic tree**.

**denominator** the lower portion of a fraction; used in calculating a ratio, proportion, or rate. For a rate, the denominator is usually the midinterval population.

**dependent variable** see **variable, dependent**.

**descriptive epidemiology** see **epidemiology, descriptive**.

**determinant** any factor that brings about change in a health condition or in other defined characteristics (see also **cause** and **risk factor**).

**direct transmission** see **transmission, direct**.

**discrete variable** (or **data**) see **variable** (or **data**), **discrete**.

**distribution** in epidemiology, the frequency and pattern of health-related characteristics and events in a population. In statistics, the frequency and pattern of the values or categories of a variable.

**dose-response** association between an exposure and health outcome that varies in a consistently increasing or decreasing fashion as the amount of exposure (dose) increases.

**dot plot** a visual display of the specific data points of a variable.

**droplet nuclei** the residue of dried droplets of infectious agents that is easily inhaled and exhaled and can remain suspended in air for relatively long periods or be blown over great distances.

**droplet spread** the direct transmission of an infectious agent by means of the aerosols produced in sneezing, coughing, or talking that travel only a short distance before falling to the ground.

## E

**effect** the result of a cause.

**effectiveness** the ability of an intervention or program to produce the intended or expected results in the field.

**efficacy** the ability of an intervention or program to produce the intended or expected results under ideal conditions.

**efficiency** the ability of an intervention or program to produce the intended or expected results

with a minimum expenditure of time and resources.

**EIS** Epidemic Intelligence Service; CDC's 2-year training program in applied epidemiology for public health professionals (<http://www.cdc.gov/eis/>).

**endemic** the constant presence of an agent or health condition within a given geographic area or population; can also refer to the usual prevalence of an agent or condition.

**environmental factor** an extrinsic factor (e.g., geology, climate, insects, sanitation, or health services) that affects an agent and the opportunity for exposure.

**epidemic** the occurrence of more cases of disease, injury, or other health condition than expected in a given area or among a specific group of persons during a particular period. Usually, the cases are presumed to have a common cause or to be related to one another in some way (see also **outbreak**).

**epidemic curve** a histogram that displays the course of an outbreak or epidemic by plotting the number of cases according to time of onset.

**epidemic period** the time span of an outbreak or epidemic.

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

**epidemiology** the study of the distribution and determinants of health conditions or events among populations and the application of that study to control health problems.

**epidemiology, analytic** the aspect of epidemiology concerned with why and how a health problem occurs. Analytic epidemiology uses comparison groups to provide baseline or expected values so that associations between exposures and outcomes can be quantified and hypotheses about the cause of the problem can be tested (see also **study, analytic**).

**epidemiology, applied** the application or practice of epidemiology to control and prevent health problems.

**epidemiology, descriptive** the aspect of epidemiology concerned with organizing and summarizing data regarding the persons affected (e.g., the characteristics of those who became ill), time (e.g., when they become ill), and place (e.g., where they might have been exposed to the cause of illness).

**epidemiology, field** applied epidemiology (i.e., the application or practice of epidemiology to control and prevent health problems), particularly when the epidemiologist(s) must travel to and work in the community in which the health problem is occurring or has occurred.

**evaluation** systematic and objective examination of activities to determine their relevance, effectiveness, and impact.

**excess risk** risk difference, calculated as the risk among the exposed group minus the risk among the unexposed group.

**experimental study** see **study, experimental**.

**exposed group** a group whose members have had contact with a suspected cause of, or possess a characteristic that is a suspected determinant of, a particular health problem.

**exposure** having come into contact with a cause of, or possessing a characteristic that is a determinant of, a particular health problem.

## F

**false-negative** a negative test result for a person who actually has the condition similarly, a person who has the disease (perhaps mild or variant) but who does not fit the case definition, or a patient or outbreak not detected by a surveillance system.

**false-positive** a positive test result for a person who actually does not have the condition. Similarly, a person who does not have the disease but who nonetheless fits the case definition, or a patient or outbreak erroneously identified by a surveillance system.

**field epidemiology** see **epidemiology, field**.

**follow-up study** see **study, cohort**.

**fomite** an inanimate object that can be the vehicle for transmission of an infectious agent (e.g., bedding, towels, or surgical instruments).

**forest plot** a graph that displays the point estimates and confidence intervals of individual studies included in a meta-analysis or systematic review as a series of parallel lines.

**frequency** the amount or number of occurrences of an attribute or health outcome among a population.

**frequency distribution** a complete summary of the frequencies of the values or categories of a variable, often displayed in a two-column table with the individual values or categories in the left column and the number of observations in each category in the right column.

**frequency polygon** a graph of a frequency distribution in which values of the variable are plotted on the horizontal axis, and the number of observations are plotted on the vertical axis. Data points are plotted at the midpoints of the intervals and are connected with straight lines.

## G

**geometric mean** see **mean, geometric**.

**graph** a visual display of quantitative data arranged on a system of coordinates.

## H

**health** a state of complete physical, mental, and social well-being and not merely the absence of

disease or other infirmity.

**health indicator** any of a variety of measures (e.g., mortality rate) that indicate the state of health of a given population.

**health information system** a combination of health statistics from different sources. Data from these systems are used to learn about health status, health care, provision and use of services, and the impact of services and programs on health.

**healthy worker effect** the observation that employed persons generally have lower mortality rates than the general population, because persons with severe, disabling disease (who have higher mortality rates) tend to be excluded from the workforce.

**herd immunity** see **immunity, herd**.

**high-risk group** a group of persons whose risk for a particular disease, injury, or other health condition is greater than that of the rest of their community or population.

**HIPAA** the Health Insurance Portability and Accountability Act, enacted in 1996, which addresses the privacy of a person's medical information as well as postemployment insurance and other health-related concerns.

**histogram** a visual representation of the frequency distribution of a continuous variable. The class intervals of the variable are grouped on a linear scale on the horizontal axis, and the class frequencies are grouped on the vertical axis. Columns are drawn so that their bases equal the class intervals (i.e., so that columns of adjacent intervals touch), and their heights correspond to the class frequencies.

**host** a person or other living organism that is susceptible to or harbors an infectious agent under natural conditions.

**host factor** an intrinsic factor (e.g., age, race/ethnicity, sex, or behaviors) that influences a person's exposure, susceptibility, or response to an agent.

**hyperendemic** the constant presence at high incidence and prevalence of an agent or health condition within a given geographic area or population.

**hypothesis** a supposition, arrived at from observation or reflection, that leads to refutable predictions; any conjecture cast in a form that will allow it to be tested and refuted.

**hypothesis, alternative** the supposition that an exposure is associated with the health condition under study. The alternative is adopted if the null hypothesis (see also **hypothesis, null**) proves implausible.

**hypothesis, null** the supposition that two (or more) groups do not differ in the measure of interest (e.g., incidence or proportion exposed); the supposition that an exposure is not associated with the health condition under study, so that the risk ratio or odds ratio equals 1. The null hypothesis is used in conjunction with statistical testing.

I

**immunity, active** resistance developed in response to an antigen (i.e., an infecting agent or vaccine), usually characterized by the presence of antibody produced by the host.

**immunity, herd** the resistance to an infectious agent of an entire group or community (and, in particular, protection of susceptible persons) as a result of a substantial proportion of the population being immune to the agent. Herd immunity is based on having a substantial number of immune persons, thereby reducing the likelihood that an infected person will come in contact with a susceptible one among human populations, also called **community immunity**.

**immunity, passive** immunity conferred by an antibody produced in another host This type of immunity can be acquired naturally by an infant from its mother or artificially by administration of an antibody-containing preparation (e.g., antiserum or immune globulin).

**incidence** a measure of the frequency with which new cases of illness, injury, or other health condition occurs among a population during a specified period.

**incidence proportion** the fraction of persons with new cases of illness, injury, or other health condition during a specified period, calculated as the number of new cases divided by the size of the population at the start of the study period (see also **attack rate**).

**incidence rate** a measure of the frequency with which new cases of illness, injury, or other health condition occur, expressed explicitly per a time frame. Incidence rate is calculated as the number of new cases over a specified period divided either by the average population (usually mid-period) or by the cumulative person-time the population was at risk.

**incubation period** the time interval from exposure to an infectious agent to the onset of symptoms of an infectious disease.

**independent variable** see **variable, independent**.

**index case** see **case, index**.

**indirect transmission** see **transmission, indirect**.

**individual data** values or observations from each record (also called raw data).

**infant mortality rate** see **mortality rate, infant**.

**infection** invasion of the body tissues of a host by an infectious agent, whether or not it causes disease.

**infectivity** the ability of an infectious agent to cause infection, measured as the proportion of persons exposed to an infectious agent who become infected.

**information bias** see **bias, information**.

**interquartile range** a measure of spread representing the middle 50% of the observations, calculated as the difference between the third quartile (75<sup>th</sup> percentile) and the first quartile (25<sup>th</sup>

percentile).

**isolation** the separation of infected persons to prevent transmission to susceptible ones. Isolation refers to separation of ill persons; **quarantine** refers to separation of potentially exposed but well persons.

## L

**latency period** the time from exposure to a causal agent to onset of symptoms of a (usually noninfectious) disease (see also **incubation period**).

**life expectancy** a statistical projection of the average number of years a person of a given age is expected to live, if current mortality rates continue to apply.

**line graph, arithmetic-scale** a graph that displays patterns or trends by plotting the frequency (e.g., number, proportion, or rate) of a characteristic or event during some variable, usually time. The y-axis, measuring frequency, uses an arithmetic scale.

**line graph, semilogarithmic-scale** a graph that displays patterns or trends by plotting the frequency (e.g., number, proportion, or rate) of a characteristic or event during some variable, usually time. The y-axis, measuring frequency, uses a logarithmic scale.

**line listing** a type of epidemiologic database, organized similar to a spreadsheet with rows and columns in which information from cases or patients are listed each column represents a variable, and each row represents an individual case or patient.

**logarithmic transformation** conversion of nominal or ordinal data to logarithmic data. The purpose is to examine rate of change instead of amount of change only.

## M

**map, area (shaded, choropleth)** a visual display of the geographic pattern of a health problem, in which a marker is placed on a map to indicate where each affected person lives, works, or might have been exposed.

**mean (or average)** commonly called the average; it is the most common measure of central tendency.

**mean, arithmetic** the measure of central location, commonly called the average, calculated by adding all the values in a group of measurements and dividing by the number of values in the group.

**mean, geometric** the mean, or average, of a set of data measured on a logarithmic scale.

**measure of association** a quantified relationship between exposure and a particular health problem (e.g., risk ratio, rate ratio, and odds ratio).

**measure of central location** a central value that best represents a distribution of data. Common measures of central location are the mean, median, and mode also called the measure of central

tendency.

**measure of dispersion** see **measure of spread**.

**measure of spread** a measure of the distribution of observations out from its central value. Measures of spread used in epidemiology include the interquartile range, variance, and the standard deviation.

**measurement scale** the complete range of possible values for a measurement.

**mechanical transmission** see **transmission, mechanical**.

**median** the measure of central location that divides a set of data into two equal parts, above and below which lie an equal number of values (see also **measure of central location**).

**medical surveillance** see **surveillance, medical**.

**midrange** the halfway point, or midpoint, in a set of observations. For the majority of data, the midrange is calculated by adding the smallest observation and the largest observation and dividing by two. The midrange is usually calculated as an intermediate step in determining other measures.

**mode** the most frequently occurring value in a set of observations (see also **measure of central location**).

**mode of transmission** the manner in which an agent is transmitted from its reservoir to a susceptible host (see also **transmission**).

**morbidity** disease; any departure, subjective or objective, from a state of physiological or psychological health and well-being.

**mortality** death.

**mortality rate** a measure of the frequency of occurrence of death among a defined population during a specified time interval.

**mortality rate, age-adjusted** a mortality rate that has been statistically modified to eliminate the effect of different age distributions among different populations.

**mortality rate, age-specific** a mortality rate limited to a particular age group, calculated as the number of deaths among the age group divided by the number of persons in that age group, usually expressed per 100,000.

**mortality rate, cause-specific** the mortality rate from a specified cause, calculated as the number of deaths attributed to a specific cause during a specified time interval among a population divided by the size of the midinterval population.

**mortality rate, crude** a mortality rate from all causes of death for an entire population, without adjustment.

**mortality rate, infant** the mortality rate for children aged <1 year, calculated as the number of

deaths reported among this age group during a given period divided by the number of live births reported during the same period, and expressed per 1,000 live births. Infant mortality rate is a universally accepted indicator of the health of a nation's population and the adequacy of its health-care system.

**mortality rate, neonatal** the mortality rate for children from age birth up to, but not including, 28 days. In calculating neonatal mortality rates, the numerator is the number of deaths among this age group during a given period, and the denominator is the number of live births reported during the same period. The neonatal mortality rate is usually expressed per 1,000 live births.

**mortality rate, postneonatal** the mortality rate for children from age 28 days up to, but not including, 1 year. In calculating postneonatal mortality rates, the numerator is the number of deaths among this age group during a given period, and the denominator is the number of live births during the same period. The postneonatal mortality rate is usually expressed per 1,000 live births.

**mortality rate, race/ethnic-specific** a mortality rate limited to a specified racial or ethnic group both numerator and denominator are limited to that group.

**mortality rate, sex-specific** a mortality rate among either males or females.

## N

**natural history of disease** the progression of a disease process in a person from the time it begins to the time it resolves, in the absence of treatment.

**NCHS** The National Center for Health Statistics, the US governmental organization responsible for national vital statistics and multiple national health surveys. Organizationally, NCHS is a component of the Centers for Disease Control and Prevention, one of the agencies of the US Department of Health and Human Services.

**NHANES** The National Health and Nutrition Examination Survey, a representative survey of the civilian, noninstitutionalized US population conducted by the National Center for Health Statistics, designed to (1) estimate the proportion of the US population and designated groups with selected disease and risk factors; (2) monitor trends in selected behaviors, exposures, and diseases; and (3) study the associations among diet, nutrition, and health.

**necessary cause** see **cause, necessary**.

**neonatal mortality rate** see **mortality rate, neonatal**.

**nominal scale** see **scale, nominal**.

**normal curve** the bell-shaped curve that results when a normal distribution is graphed.

**normal distribution** a distribution represented as a bell shape, symmetrical on both sides of the peak, which is simultaneously the mean, median, and mode, and with both tails extending to infinity.

**notifiable disease** a disease that, by law, must be reported to public health authorities upon diagnosis.

**null hypothesis** see **hypothesis, null**.

**numerator** the upper portion of a fraction (see also **denominator**).

## O

**observational study** see **study, observational**.

**odds ratio** a measure of association used in comparative studies, particularly case-control studies, that quantifies the association between an exposure and a health outcome; also called the cross-product ratio.

**ordinal scale** see **scale, ordinal**.

**outbreak** the occurrence of more cases of disease, injury, or other health condition than expected in a given area or among a specific group of persons during a specific period. Usually, the cases are presumed to have a common cause or to be related to one another in some way. Sometimes distinguished from an epidemic as more localized, or the term less likely to evoke public panic (see also **epidemic**).

**outbreak, common-source** an outbreak that results from persons being exposed to the same harmful influence (e.g., an infectious agent or toxin). The exposure period can be brief or can extend over days, weeks, or longer, with the exposure being either intermittent or continuous.

**outbreak, point-source** a common source outbreak in which the exposure period is relatively brief so that all cases occur within one incubation period.

**outbreak, propagated** an outbreak that spreads from person to person rather than from a common source.

**outcome(s)** any or all of the possible results that can stem from exposure to a causal factor or from preventive or therapeutic interventions; all identified changes in health status that result from the handling of a health problem.

**outlier** a value substantively or statistically different from all (or approximately all) of the other values in a distribution.

## P

**P value** the probability of observing an association between two variables or a difference between two or more groups as large or larger than that observed, if the null hypothesis were true. Used in statistical testing to evaluate the plausibility of the null hypothesis (i.e., whether the observed association or difference plausibly might have occurred by chance).

**pandemic** an epidemic occurring over a widespread area (multiple countries or continents) and usually affecting a substantial proportion of the population.

**passive immunity** see **immunity, passive**.

**passive surveillance** see **surveillance, passive**.

**pathogenicity** the ability of an agent to cause disease after infection, measured as the proportion of persons infected by an agent who then experience clinical disease.

**percentile** a set of cut points used to divide a distribution or a set of ranked data into 100 parts of equal area with each interval between the points containing 1/100 or 1% of the observations. For example, the 5<sup>th</sup> percentile is a cut point with 5% of the observations below it and the remaining 95% above it.

**period prevalence** see **prevalence, period**.

**person-time rate** the incidence rate calculated as the number of new cases among a population divided by the cumulative person-time of that population, usually expressed as the number of events per persons per unit of time.

**person-time** the amount of time each participant in a cohort study is observed and disease-free, often summed to provide the denominator for a person-time rate.

**phylogenetic tree** a branching chart that indicates the evolutionary lineage or genetic relatedness of organisms.

**pie chart** a circular graph of a frequency distribution in which each segment of the pie is proportional in size to the frequency of corresponding category.

**point prevalence** see **prevalence, point**.

**point-source outbreak** see **outbreak, point-source**.

**population** the total number of inhabitants of a geographic area or the total number of persons in a particular group (e.g., the number of persons engaged in a certain occupation).

**population pyramid** a graphical display of the age-sex distribution of a population, constructed with a horizontal histogram of the age distribution of males pointing to the left, and the corresponding horizontal histogram of age distribution of females pointing to the right.

**portal of entry** a pathway into the host that gives an agent access to tissue that will allow it to multiply or act.

**portal of exit** a pathway by which an agent can leave its host.

**postneonatal mortality rate** see **mortality rate, postneonatal**.

**predictive value positive** the proportion of cases identified by a test, reported by a surveillance system, or classified by a case definition that are true cases, calculated as the number of true-positives divided by the number of true-positives plus false-positives.

**prevalence** the number or proportion of cases or events or attributes among a given population.

**prevalence rate** the proportion of a population that has a particular disease, injury, other health condition, or attribute at a specified point in time (point prevalence) or during a specified period (period prevalence).

**prevalence, period** the amount of a particular disease, chronic condition, or type of injury present among a population at any time during a particular period.

**prevalence, point** the amount of a particular disease, chronic condition, or type of injury present among a population at a single point in time.

**privacy rule** a set of regulations based on the Health Insurance Portability and Accountability Act to protect the privacy of individually identifiable health information.

**propagated outbreak** see **outbreak, propagated**.

**proportion** a ratio in which the numerator is included in the denominator; the ratio of a part to the whole, expressed as a "decimal fraction" (e.g., 0.2), a fraction (1/5), or a percentage (20%).

**proportion, attributable** a measure of the impact of a causative factor on the public health; the proportion of a health state or event among exposed persons that can be attributed to the exposure also called attributable risk percent.

**proportionate mortality** the proportion of deaths among a population attributable to a particular cause during a selected period. Each cause of death is expressed as a percentage of all deaths, and the sum of the proportionate mortality for all causes must equal 100%. These proportions are not mortality rates because, in proportionate mortality, the denominator is all deaths instead of the population among whom the deaths occurred.

**prospective study** see **study, prospective**.

## Q

**quarantine** the separation of well persons who have been exposed or are suspected to have been exposed to a communicable disease, to monitor for illness and to prevent potential transmission of infection to susceptible persons during the incubation period. Quarantine refers to separation of potentially exposed but well persons; **isolation** refers to separation of ill persons.

## R

**race/ethnic-specific mortality rate** see **mortality rate, race/ethnic-specific**.

**random sample** see **sample, random**.

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

**rate** an expression of the relative frequency with which an event occurs among a defined population per unit of time, calculated as the number of new cases or deaths during a specified period divided by either person-time or the average (midinterval) population. In epidemiology, it

is often used more casually to refer to proportions that are not truly rates (e.g., attack rate or case-fatality rate).

**rate ratio** a measure of association that quantifies the relation between an exposure and a health outcome from an epidemiologic study, calculated as the ratio of incidence rates or mortality rates of two groups.

**ratio** the relative size of two quantities, calculated by dividing one quantity by the other.

**record** in a line listing, each row is a record or observation. A record represents data related to a single case.

**relative risk** a general term for measures of association calculated from the data in a two-by-two table, including risk ratio, rate ratio, and odds ratio (see also **risk ratio**).

**representative sample** see **sample, representative**.

**reservoir** the habitat in which an infectious agent normally lives, grows, and multiplies, which can include humans, animals, or the environment.

**retrospective study** see **study, retrospective**.

**risk** the probability that an event will occur (e.g., that a person will be affected by, or die from, an illness, injury, or other health condition within a specified time or age span).

**risk factor** an aspect of personal behavior or lifestyle, an environmental exposure, or a hereditary characteristic that is associated with an increase in the occurrence of a particular disease, injury, or other health condition.

**risk ratio** a measure of association that quantifies the association between an exposure and a health outcome from an epidemiologic study, calculated as the ratio of incidence proportions of two groups.

## S

**sample** a selected subset of a population a sample can be random or nonrandom and representative or nonrepresentative.

**sample, random** a sample of persons chosen in such a way that each one has the same (and known) probability of being selected.

**sample, representative** a sample whose characteristics correspond to those of the original or reference population.

**scale, interval** a measurement scale consisting of quantitative categories whose values are measured on a scale of equally spaced units, but without a true zero point (e.g., date of birth).

**scale, nominal** a measurement scale consisting of qualitative categories whose values have no inherent statistical order or rank (e.g., categories of race/ethnicity, religion, or country of birth).

**scale, ordinal** a measurement scale consisting of qualitative categories whose values have a distinct order but no numerical distance between their possible values (e.g., stage of cancer, I, II, III, or IV).

**scale, ratio** a measurement scale consisting of quantitative categories whose values are intervals with a true zero point (e.g., height in centimeters or duration of illness).

**scatter diagram** (or **scattergram**) a graphical display of the association between two variables in which a dot is plotted on the graph for each set of paired values for two continuous variables, with one variable plotted on the horizontal axis, and the other plotted on the vertical axis.

**seasonality** change in physiologic status or in the occurrence of a disease, chronic condition, or type of injury that conforms to a regular seasonal pattern.

**secondary attack rate** see **attack rate, secondary**.

**secular trend** see **trend, secular**.

**selection bias** see **bias, selection**.

**semilogarithmic-scale line graph** see **line graph, semilogarithmic-scale**

**sensitivity** the ability of a test, case definition, or surveillance system to identify true cases; the proportion of people with a health condition (or the proportion of outbreaks) that are identified by a screening test or case definition (or surveillance system).

**sentinel surveillance** see **surveillance, sentinel**.

**sex-specific mortality rate** see **mortality rate, sex-specific**.

**skewed** a distribution that is not symmetrical.

**source (of infection)** the person, animal, object, or substance from which an infectious agent is transmitted to a host.

**source case** see **case, source**.

**specificity** the ability of a test, case definition, or surveillance system to exclude persons without the health condition of interest; the proportion of persons without a health condition that are correctly identified as such by a screening test, case definition, or surveillance system.

**spectrum of illness** the range of manifestations a disease process can take (e.g., from asymptomatic to mild clinical illness to severe illness and death).

**sporadic** an event that occurs infrequently and irregularly.

**spot map** a visual display of the geographic pattern of a health problem, in which a marker is placed on a map to indicate where each affected person lives, works, or might have been exposed.

**standard deviation** a statistical summary of how dispersed the values of a variable are around

its mean, calculated as the square root of the variance.

**standard error (of the mean)** the standard deviation of a theoretical distribution of sample means of a variable around the true population mean of that variable. Standard error is computed as the standard deviation of the variable divided by the square root of the sample size.

**statistical inference** generalizations developed from sample data, usually with calculated degrees of uncertainty.

**statistical significance** the measure of how likely it is that a set of study results could have occurred by chance alone. Statistical significance is based on an estimate of the probability of the observed or a greater degree of association between independent and dependent variables occurring under the null hypothesis (see also **P value**).

**study, analytic** a study, usually observational, in which groups are compared to identify and quantify associations, test hypotheses, and identify causes. Two common types are cohort studies and case-control studies.

**study, case-control** an observational analytic study that enrolls one group of persons with a certain disease, chronic condition, or type of injury (case-patients) and a group of persons without the health problem (control subjects) and compares differences in exposures, behaviors, and other characteristics to identify and quantify associations, test hypotheses, and identify causes.

**study, cohort** an observational analytic study in which enrollment is based on status of exposure to a certain factor or membership in a certain group. Populations are followed, and disease, death, or other health-related outcomes are documented and compared. Cohort studies can be either prospective or retrospective.

**study, cross-sectional** a study in which a sample of persons from a population are enrolled and their exposures and health outcomes are measured simultaneously; a survey.

**study, experimental** a study in which the investigator specifies the type of exposure for each person (clinical trial) or community (community trial) then follows the persons' or communities' health status to determine the effects of the exposure.

**study, observational** a study in which the investigator observes rather than influences exposure and disease among participants. Case-control and cohort studies are observational studies (see also **study, experimental**).

**study, prospective** an analytic study in which participants are enrolled before the health outcome of interest has occurred.

**study, retrospective** an analytic study in which participants are enrolled after the health outcome of interest has occurred. Case-control studies are inherently retrospective.

**subclinical** without apparent symptoms.

**surveillance, active** public health surveillance in which the health agency solicits reports.

**surveillance, medical** monitoring of a person who might have been exposed to an infectious, chemical, radiologic, or other potentially causal agent, for the purpose of detecting early symptoms.

**surveillance, passive** public health surveillance in which data are sent to the health agency without prompting.

**surveillance, sentinel** a surveillance system that uses a prearranged sample of sources (e.g., physicians, hospitals, or clinics) who have agreed to report all cases of one or more notifiable diseases.

**surveillance, syndromic** (1) the monitoring of the frequency of illnesses with a specified set of clinical features among a given population without regard to the specific diagnoses, if any, that are assigned to them by clinicians. (2) A system for early detection of outbreaks whereby health department staff, assisted by automated acquisition of data routinely collected for other purposes and computer generation of statistical signals, monitor disease indicators, particularly those associated with possible terrorism-related biologic and chemical agents, continually or at least daily to detect outbreaks earlier than would otherwise be possible with traditional public health methods.

**survey** a systematic canvassing of persons to collect information, often from a representative sample of the population.

**survival curve** a line graph that begins with 100% of the study population and displays the percentage of the population still surviving at successive points in time. A survival curve can also be used to depict freedom from a health problem, complication, or another endpoint.

**symmetrical** a type of distribution where the shapes to the right and left of the central location are the same. Normal, bell-shaped distributions are symmetrical; the mean, median, and mode are the same.

**symptom** any indication of disease noticed or felt by a patient.

**syndrome** a combination of symptoms characteristic of a disease or health condition; sometimes refers to a health condition without a clear cause (e.g., chronic fatigue syndrome).

**syndromic surveillance** see **surveillance, syndromic**.

## T

**table** an arrangement of data in rows and columns. In epidemiology, the data are usually summaries of the frequency of occurrence of an event or characteristic occurring among different groups.

**table shell** a table that is completely drawn and labeled but contains no data.

**table, two-by-two** a two-variable table with cross-tabulated data, in which each variable has only two categories. Usually, one variable represents a health outcome, and one represents an exposure or personal characteristic.

**transmission (of infection)** any mode or mechanism by which an infectious agent is spread to a susceptible host.

**transmission, airborne** transfer of an agent suspended in the air, considered a type of indirect transmission.

**transmission, biologic** indirect transmission by a vector in which the infectious agent undergoes biologic changes inside the vector as part of its life cycle before it is transmitted to the host (see also **transmission, mechanical**).

**transmission, direct** immediate transfer of an agent from a reservoir to a host by direct contact or droplet spread.

**transmission, indirect** transfer of an agent from a reservoir to a host either by being suspended in air particles (airborne), carried by an inanimate objects (vehicleborne), or carried by an animate intermediary (vectorborne).

**transmission, mechanical** indirect transmission by a vector in which the infectious agent does not undergo physiologic changes inside the vector (see also **transmission, biologic**).

**transmission, vectorborne** transmission of an agent by a living intermediary (e.g., tick, mosquito, or flea); considered a type of indirect transmission.

**transmission, vehicleborne** transmission of an agent by an inanimate object; considered a type of indirect transmission; includes foodborne and waterborne transmission.

**trend** movement or change in frequency over time, usually upwards or downwards.

**trend, secular** changes occurring over a substantial period, generally years or decades.

**trial, clinical** an experimental study that uses data from individual persons. The investigator specifies the type of exposure for each study participant and then follows each person's health status to determine the effects of the exposure.

**trial, community** an experimental study that uses data from communities. The investigator specifies the type of exposure for each community and then follows the communities' health status to determine the effects of the exposure.

**trial, randomized clinical** a clinical trial in which persons are randomly assigned to exposure or treatment groups.

**two-by-two table** see **table, two-by-two**.

## V

**validity** the degree to which a measurement, questionnaire, test, or study or any other data-collection tool measures what it is intended to measure.

**variable** any characteristic or attribute that can be measured and can have different values.

**variable (or data), discrete** a variable that is limited to a finite number of values; data for such a variable.

**variable, continuous** a variable that has the potential for having an infinite number of values along a continuum (e.g., height and weight).

**variable, dependent** in a statistical analysis, a variable whose values are a function of one or more other variables.

**variable, independent** an exposure, risk factor, or other characteristic being observed or measured that is hypothesized to influence an event or manifestation (the dependent variable).

**variance** a measure of the spread in a set of observations, calculated as the sum of the squares of deviations from the mean, divided by the number of observations minus 1 (see also **standard deviation**).

**vector** a living intermediary that carries an agent from a reservoir to a susceptible host (see also **transmission, biologic** and **transmission, mechanical**) (e.g., mosquitoes, fleas, or ticks).

**vehicle** an inanimate object that can carry an agent from a reservoir to a susceptible host (e.g., food, water, blood products, and bedding) (see also **transmission, indirect**).

**virulence** the ability of an infectious agent to cause severe disease, measured as the proportion of persons with the disease who become severely ill or die.

**vital statistics** systematically tabulated data about recorded births, marriages, divorces, and deaths.

## X

**x-axis** the horizontal axis of a rectangular graph, usually displaying the independent variable (e.g., time).

## Y

**y-axis** the vertical axis of a rectangular graph, usually displaying the dependent variable (e.g., frequency — number, proportion, or rate).

**years of potential life lost (YPLL)** a measure of the impact of premature death on a population, calculated as the sum of the differences between a predetermined minimally acceptable age (e.g., 65 years or current life expectancy) and the age at death for everyone who died earlier than that age.

## Z

**zoonosis** an infectious disease that is transmissible from animals to humans.