

A SHORT GLOSSARY OF EPIDEMIOLOGIC TERMS

The following list of terms was compiled to assist those unfamiliar with the language of epidemiology. The terms come from various sources and an attempt has been made to simplify definitions where possible. As a result they may suffer from a lack of rigor but will suffice for normal usage. The list has been tested on persons both familiar and unfamiliar with veterinary epidemiology. The list is purposely brief and does not contain terms which are generally self evident.

Accuracy - the degree to which a measurement, or an estimate based on measurements, represents the true value of the attribute that is being measured.

Agent - a factor such as a microorganism or chemical substance whose presence or excessive presence is necessary for the occurrence of a disease.

Analysis of variance - also called ANOVA; a statistical technique that isolates and assesses the contribution of categorical variables, each with more than 2 classes, to variation in the mean of a continuous variable. See also t-test.

Analytic study - a hypothesis testing method of investigating the association between a given disease, health state, or other outcome variable, and possible causative factors.

Association - the degree of statistical dependence between two or more events or variables. Events are said to be associated when they occur more frequently together than one would expect by chance. Association does not necessarily imply a causal relationship. Statistical significance testing enables us to determine how unlikely it would be to observe the sample relationship by chance if in fact no relationship exists in the population that was sampled.

Attributable Risk - the excess risk (above background) that is explained by the characteristic or risk factor under study. It requires calculation of incidence rates.

Attributable risk = Incidence rate_{exp} - Incidence rate_{non-exp}

Attack Rate - the proportion of a specific population affected during an outbreak. The population is usually limited to susceptible animals or those identifiably at risk. It is a special form of cumulative incidence that is used in an outbreak investigation.

Benefit-Cost Ratio - the ratio of the net present values (usually monetary values) of measurable benefits to costs. Used to determine the economic feasibility or probability of success of a time-bounded program.

Bias - any effect at any stage of an investigation tending to produce results that depart systematically from the true values i.e. a systematic error.

Bias (Response bias) - a systematic error due to differences in characteristics between those who volunteer to participate in a study and those who do not.

Bias (Selection bias) - error due to systematic differences in characteristics between those animals or herds which are selected for study and those which are not.

Bimodal Distribution - a distribution with two regions of high frequency separated by region of low frequency of observation.

Binomial Distribution - a probability distribution associated with two mutually exclusive outcomes such as yes or no.

Case-Control Study - a study that starts with the identification of animals (or herds) with the disease of interest and a suitable control (comparison, reference) group of animals (or herds) without the disease. This type of study involves collection and analysis of data on disease determinants in the two groups. Usually, it is a retrospective study because disease events have occurred before the exposure history is determined.

Case-Fatality Rate - the proportion of animals contracting a disease that die of that disease during a specified follow-up period.

Case-fatality rate = $\frac{\text{No. of deaths from specific cause}}{\text{No. cases of specific cause}}$

Categorical Data - qualitative data which can be allocated to specific groups. May be nominal (ie. named) or ordinal (ie. ordered) or dichotomous (ie. presence/absence).

Cause, Necessary - a variable which must always precede an effect. This effect need not be a sole result of the one variable.

Chi-Square Test - a method of testing to determine whether two or more series of proportions or frequencies are significantly different from one another or whether a single series of proportions differs significantly from an expected distribution. Pearson's Chi-square is used for unmatched data and McNemar's Chi-square for matched data. See definition of association for further explanation.

Clustering - a closely grouped series of events or cases of a disease in relation to time or place or both. The term is normally used to describe aggregation of relatively uncommon events or diseases.

Cohort Study - An epidemiologic study in which subsets of a defined population are sampled on the basis of exposure to a factor or factors hypothesized to influence the probability of occurrence of a given disease or other outcome. It is commonly undertaken prospectively.

Confidence Limits - an interval whose end points can be calculated from observational data and has a specified probability of containing the parameter of interest.

Confounding - a situation in which the effects of two factors are not separated. The distortion of the apparent effect of an exposure or risk factor brought about by association with other factors that can influence the outcome.

Confounding Factor - a confounding factor or variable is one which is distributed non-randomly with respect to the independent (exposure) variable and is associated with the dependent (outcome) variable being studied. The association with the dependent variable is usually established from results of previous studies.

Contingency Table - a tabular cross-classification of data

such that subcategories of one characteristic are indicated horizontally (in rows) and subcategories of another characteristic are indicated vertically (in columns), and the number of units in each cell is indicated. The simplest contingency table is the fourfold or 2 x 2 table, but a contingency table may include several dimensions of classification.

Continuous Data - quantitative data with a potentially infinite number of possible values along a continuum.

Correlation Coefficient - a measure of association that indicates the degree to which two or more sets of observations fit a linear relationship. This coefficient, represented by the letter "r", can vary between +1 and -1. If "r" is +1, there is a perfect linear relationship in which one variable varies directly with the other. If "r" = -1, there is again a perfect linear association but one variable varies inversely with the other.

Cost Benefit Analysis - methods of identifying the losses and gains in monetary terms of the effects of a disease that are incurred by society as a whole.

Cross-Sectional Study - (syn: prevalence study) - a study carried out on a representative sample of a population that examines the relationship between a disease or other health-related characteristic and other variables of interest as they exist in a defined population at one particular time.

Crude Rate - a rate which applies to a total population irrespective of the attributes of that population (cf. specific rate).

Data - facts of any kind. Data are plural, datum is singular.

Data Base - a systemized collection of information, commonly on electronic media about a specific subject such as animal disease.

Decision Analysis - application of probability theory with the aim of calculating the optimal strategy from a series of alternative decisions which are often expressed graphically in the form of a decision tree. Decision analysis is a tool to help stock owners decide which of several options eg vaccination or culling is the optimal alternative for treatment or control of a disease.

Degrees of Freedom (df) - in a contingency table is the number of independent comparisons that can be made between the members of a sample. It is one less than the number of row categories multiplied by one less than the number of column categories.

Denominator - the population at risk in the calculation of a rate or ratio. See also Numerator

Dependent Variable - (syn: outcome/response variable) a variable or factor, the value of which depends on or is hypothesized to depend on the effect of other [causal] variable(s) in the study.

Determinant - any factor whether event, characteristic or other definable entity, that brings about change in a health condition or other defined characteristic.

Deterministic Model - a mathematical model which assumes all parameters and variables are constant and not random variables.

Disease Determinant - any variable (factor) associated with a disease which if removed or altered results in a change in the incidence of disease in a population.

Disease Reservoir - any animal or object in which an infectious agent multiplies or develops and upon which it depends as a species for survival in nature.

Discriminant Analysis - a statistical technique similar to regression analysis but where the response, or dependent variable is dichotomous. Alternatively - a statistical method used to allocate an individual to one or more distinct groups.

Distribution-free Method (syn: nonparametric method) - a method of testing a hypothesis or of setting up a confidence interval that does not depend on the form of the underlying distribution; in particular it does not depend upon the variable following a normal distribution.

Dose-Response Relationship - a relationship in which change in amount, intensity or duration of exposure to a factor is

associated with a change (either an increase or a decrease) in risk of a specified outcome.

Endemic Disease - the constant presence of a disease or infectious agent within a given geographic area or population group. It also implies a prevalence which is usual in the area or in the population.

Epidemic - the occurrence in a population or region of cases of disease clearly in excess of normal expectancy - this is frequently taken as more than two standard deviations greater than the mean occurrence.

Epidemic curve - a histogram in which the X-axis represents the time of occurrence of disease cases and the Y-axis represents the frequency of disease cases. It is a useful tool to determine the epidemiology of disease occurrence in an outbreak investigation.

Epidemic-Propagating - an outbreak or series of outbreaks resulting from animal to animal spread.

Epidemiology - the study of the distribution and determinants of health related states and events in populations. It is a term now in common usage for studies in animal populations although epizootiology is still occasionally used.

Epidemiology, Descriptive - study of the occurrence of disease or other health related characteristics in populations. Implies general observation rather than analysis.

Error - Sampling - after testing a sample from a large population, the mean or any other statistic calculated from the sample will have a different value from the true value if the whole population was measured. The difference between the value for the whole population and its estimate calculated from the sample is called the sampling error.

Error-Systematic - that due to factors other than chance, such as faulty measuring instruments.

Experimental Epidemiology - the planning of specific population experiments to test epidemiological hypotheses (eg field trials, clinical trials).

Experimental Study - a study in which the conditions are under the direct control of the investigator.

Factor Analysis - a set of statistical methods for analyzing the correlations among several variables to estimate the number of fundamental dimensions that underlie the observed data and to describe and measure those dimensions. Used frequently in the development of a scoring system for rating scales and questionnaires.

Factorial Design - a method of setting up an experiment or study to assure that all levels of each controlled variable occur with all levels of the others.

False Negative - when the result of an individual test is negative but the disease or condition is present.

False Positive - when the result of an individual test is positive but the disease or condition is not present.

Frequency - a count, or number of occurrences, of an event in a specified population and time period.

Frequency Distribution - any arrangement of numerical data obtained by measuring a parameter in a population.

Histogram - frequency distribution plotted in the form of rectangles whose bases are equal to the class width and whose areas are proportional to the absolute or relative frequencies.

Hypotheses - a proposition that can be tested by facts that are known or can be obtained. The assertion that an association between two, or more variables or a difference between 2 or more groups, exists in the larger population of interest.

Incidence - the number of new cases of disease or other condition which occur in a specified population during a given period.

Mathematically, 2 types of incidence rate can be distinguished. These are incidence density rates and cumulative incidence.

Incubation Period - the interval of time between invasion by an infectious agent or contact with a chemical and the

appearance of symptoms of the disease or condition in question.

Independent Variable - the characteristic being observed or measured that is hypothesized to influence an event. An independent variable is not influenced by the event or manifestation but may cause it or contribute to its variation.

Index Case - the first diagnosed case of an outbreak in a herd or other defined group.

Infectivity - the ability of an agent to enter, survive and multiply in the host.

Inference - the process of passing from observations to generalizations.

Intervention Study - an epidemiologic investigation designed to test a hypothesized cause-effect relationship by modifying a supposed causal factor in a population and measuring the change in the parameter.

Latent Infection - persistence of an infectious agent within the host without symptoms of disease.

Linear Model - a statistical model of a dependent variable Y as a linear combination of other variables (X's). Of the general form, $Y = a + bX + E$ where E represents random variation.

Linear Regression - statistical method used to study the relationship between independent and dependent variables when the dependent variable consists of continuous data.

Log-Linear Model - a statistical model that uses an analysis of variance type of approach for the modelling of frequency counts in contingency tables.

Longitudinal Study - a study conducted over a defined period of time which may be either retrospective or prospective. See also Case Control and Cohort Study.

Marginal cost (of animal health) - the cost of an additional amount of animal health care.

Marginal Return (of animal health) - the income obtained from using an additional amount of animal health care.

Matching - the process of making a study group and a comparison group comparable with respect to factors which are likely to influence the results but in which the experimenter has no immediate interest.

Mathematical Model - a representation of a system, process, or relationship in mathematical form in which equations are used to simulate the behavior of the system or process under study.

Mean-Arithmetic - a measure of central tendency computed by adding all the individual values together and dividing by the number in the group.

Mean-Geometric - a measure of central tendency calculable only for positive values and computed by taking the logarithms of the values, calculating their arithmetic mean and then converting back to the original units by taking the antilogarithms.

Median - the median is the middle value of a set of observations arranged in order of magnitude.

Mode - the mode is the most frequently occurring value in a set of observations. A given set of observations can have more than one mode. (see also Bimodal Distribution).

Model - a representation or simulation of an actual situation.

Monitoring - the performance and analysis of routine measurements aimed at the early detection of changes in the prevalence or incidence of disease, health, or alteration in a production parameter.

Multiple Regression - an analytical method which determines the relationship between a dependent variable and two or more independent variables.

Multistage Sampling - a term applied to the selection of a sample in two or more stages. eg, selecting a sample of herds and then a sample of livestock within those herds.

Multivariate Analysis - a set of techniques used when the

variation in several variables has to be studied simultaneously. In statistics, any analytic method that allows the simultaneous study of two or more dependent variables.

Necessary Cause - the characteristic referred to as the "cause" which is always found in the presence of the effect.

Nominal Data - a type of data in which there are limited categories but no order, such as breed and eye color.

Non-Parametric - see Distribution-Free Method.

Normal - within the usual range of variation in a given population or population group; or frequently occurring in a given population or group.

Normal Distribution - a continuous symmetrical frequency distribution where both tails extend to infinity, the arithmetic mean, mode and median are identical. Graphically it is a bell shaped curve and its steepness or shape is completely determined by the mean and variance.

Null Hypothesis - the hypothesis that two variables have no association at all, or two or more population distributions do not differ from each other.

Numerator - the upper portion of a fraction used to calculate a rate or ratio.

Observational Study - an epidemiological study where nature is allowed to take its course while changes or differences in one characteristic are studied in relation to changes or differences in other(s) without intervention of the investigator (e.g. descriptive, cross-sectional case-control, cohort).

Occurrence - a statement indicating the presence of disease without signifying the frequency. This definition describes the use of the word in international animal disease reports.

Odds Ratio - the ratio of two odds. In a case-control study, the ratio of the odds of exposure among the cases to the odds of exposure among the non-cases. In a cohort or cross sectional study, the ratio of the odds of disease in the exposed to the odds of disease in the unexposed.

Mathematically, the odds ratio is calculated identically for all types of study design.

	Diseased	No disease
Exposed	a	b
Unexposed	c	d

The odds ratio is ad/bc .

Ordinal data - a type of data in which there are limited categories with an inherent ranking from lowest to highest (such as severity of disease).

Outbreak - the occurrence of disease in a herd or any other identifiable group of animals. For practical purposes, the term is synonymous with epidemic.

Outliers - observations differing so widely from the rest of the data as to lead one to suspect that a gross error in recording may have been committed, or suggesting that these values came from a different population.

Pandemic - an epidemic occurring over a very wide area, involving many countries and usually affecting a large proportion of the population.

Parameter - a summary descriptive characteristic of a population (cf statistic - which is a sample-based measure).

Parametric Method - a method of hypothesis testing that requires the assumption of a particular model for the distribution of data (e.g. normal, Poisson, etc.).

Path Analysis - a statistical technique for estimating the sizes of influences of one variable on another in a hypothetical chain of causation.

Pathogenicity - the ability of an organism to produce disease.

Power - probability of finding a difference between two or

more groups given that a difference exists. Power = 1-Beta = 1-Probability of a type II error.

Precision - the quality of being sharply defined or stated. Refers to the ability of a test or measuring device to give consistent results when applied repeatedly. Sometimes also called repeatability.

Predictive Value - in screening or diagnostic tests, the predictive value of a positive test is the proportion of test positive animals that have the disease. The predictive value of a negative test is the probability that an animal with a negative test does not have the disease. The predictive value of a test is determined by the sensitivity and specificity of the test, and by the prevalence of the condition at the time the test is used.

Prevalence - the proportion of cases of a disease or other condition present in a population without any distinction between old and new cases. When used without qualification the term usually refers to the number of cases as a proportion of the population at risk at a specified point in time (point prevalence).

$$\text{Prevalence} = \frac{\text{No. cases at specific point in time}}{\text{Population at risk at same point in time}}$$

Prevalence study - see cross-sectional study

Primary Case - the individual that introduces disease into a herd, flock, or other group under study. Not necessarily the first diagnosed case in that herd. See index case.

Proportion - a fraction where the numerator is a subset of the denominator.

Prospective Study - see Cohort Study.

Qualitative data - that which possess specific qualities such as breed, gender, or color. See nominal data.

Random - governed by chance.

Randomization - allocation of individuals to groups by chance. Within the limits of chance variation, randomization should make control and experimental groups similar at the start of an investigation and ensure that personal judgement and prejudices of the investigator do not influence

allocation. Note that random allocation follows a predetermined plan often devised with the aid of a table of random numbers or by an electronic random number generator.

Random Sample - a sample of a population assembled so that each member of the population has an equal and non-zero opportunity to be selected.

Random Sampling - procedure for selecting individuals from a population so that each has an equal chance of being selected in the sample.

Rate - an expression of the change in one quantity per unit time.

It is a ratio whose essential characteristic is that time is an element of the denominator and in which there is a distinct relationship between numerator and denominator. See also ratio and proportion.

Ratio - the expression of the relationship between a numerator and denominator where the two are separate and distinct quantities, i.e the numerator is not included in the denominator.

Regression Analysis - a statistical technique used to examine the relationship between two continuous variables. See also linear regression.

Relative Risk - the ratio of the disease incidence in individuals exposed to a hypothesized factor to the incidence in individuals not exposed; a measure of association commonly used in cohort studies. See also odds ratio.

	Diseased	No disease
Exposed	a	b
Unexposed	c	d

The relative risk is:
 $(a/a+b) / (c/c+d)$

Repeatability - the ability of a test to give consistent results in repeated tests. See precision.

Replication - the execution of an experiment or survey more than once to confirm the findings and obtain an improved estimate of sampling error.

Response Rate - the number of completed or returned survey instruments (questionnaires, interview etc.) divided by the total number of individuals selected for study.

Retrospective Study - a study that collects and utilizes historical data. A case-control study is retrospective because it looks back from the point of known effects to determine causative factors.

Risk Factor - an attribute or exposure that increases the probability of occurrence of disease or other specified outcome.

Robust - a statistical test is described as robust if the inferences hold true even when assumptions inherent in the tests are violated.

Sampling - the process of selecting a number of representative subjects from all the subjects in a particular group. Conclusions based on sample results may be attributed only to the population sampled. See also random sample and selection bias.

Screening - implies subjecting a population or sample of a population to a diagnostic test or procedure, with the objective of detecting disease. Tests used for this purpose are usually cheap, easily performed, sensitive but often not very specific.

Secular trend - a long term trend in the occurrence of disease or other condition.

Sensitivity - is the proportion of truly diseased animals in the screened population which are identified as diseased by the test. It is a measure of the probability that a diseased individual will be correctly identified by the test.

Sentinel Herds - herds that are reasonably representative of the population as a whole and which are tested at regular intervals for infectious disease to determine whether and to what extent the diseases are occurring in the population.

Seroepidemiology - epidemiological studies based on an

examination of sera taken from the population or a sample of the population.

Significance, Level of - also known as alpha (α) or type I error rate. The probability of saying a difference exists when none does.

Spatial distribution - the relationship of disease events to location of individual animals or clusters of animals.

Specificity - is the proportion of truly non-diseased animals correctly identified by the test. Like sensitivity, specificity is a conditional probability.

Specific Rate - expresses the frequency of a characteristic per unit size of a specific population.

Sporadic - a disease occurring irregularly and generally infrequently and without any apparent underlying pattern.

Standard Deviation - a measure of dispersion or variation. Equal to the positive square root of the variance. The mean indicates where the values for a group are centered. The standard deviation is a measure of how widely values are dispersed around the mean in the population.

Standard Error - measure of the variability of a sample statistic that specifically relates an observed mean to the true mean of the population.

Statistic - a summary value calculated from a sample of observations usually to estimate a population parameter.

Statistical Significance - statistical methods allow an estimate to be made of the probability of the observed degree of association between independent and dependent variables being exceeded under a null hypothesis. From this estimate the statistical "significance" of a result can be stated. Usually the level of statistical significance is stated by the "P" value or probability value. See also Significance, Level of.

Statistics - the science and art of dealing with variation in data through collection, classification, and appropriate analysis.

Stochastic Model - a mathematical model which takes into

consideration the presence of variability in one or more of its parameters.

Stratified Sample - involves dividing the population into distinct subgroups according to some important characteristic, eg herd size, and selecting a random sample out of each subgroup.

Surveillance - a system or measurement technique to gain knowledge about a population by collection, analysis, and interpretation of data with a view to the early detection of cases of disease or changes in the health status of the population. The goal of surveillance is directed action in the treatment or prevention of the condition.

Survey - an investigation in which information is systematically collected.

Systematic Sample - the procedure of selecting according to some simple systematic rule, such as every 5th cow in the herd as they enter the milking parlor. A systematic sample may lead to errors that invalidate generalizations.

Temporal Distribution - the relationship of disease events to time.

Theoretical Epidemiology - the development of mathematical/statistical models to explain different aspects of the occurrence of a variety of diseases.

Trend - a long-time movement in an ordered series (e.g. a time series). An essential feature is that the movement, whilst possibly irregular in the short term, shows movement consistently in the same direction over a long term.

T test - a test where data is continuous, and there are two comparison groups. It can be used to (1) estimate the mean of a normally distributed population, or to (2) test the difference between two sample means. See also analysis of variance.

Type I Error - an error which occurs when using data from a sample that demonstrates a statistically significant association when no such association is present in the population. Equals the level of significance or alpha.

Type II Error - an error that occurs from failure to demonstrate a statistically significant association when one exists in a population. Equals Beta. The power of a study equals $1 - \text{Beta}$.

Validity - the extent to which a study or test measures what it sets out to measure.

Variable - see Dependent variable, Independent variable.

Variance - the variance of a set of observations is the sum of squares of the deviation of each observation from the arithmetic mean of the observations, divided by one less than the number of observations.

Vector - a living organism (frequently an arthropod) that transports an infectious agent from an infected animal or its wastes to a susceptible individual, its food or immediate surroundings.

Virulence - it is the degree of pathogenicity and indicates the potential severity of the disease produced by an agent in a given host; numerically expressed as the ratio of the number of cases of overt infection to the total number infected, as determined by immunoassay. Sometimes, the case-fatality rate is considered an indicator for the virulence of disease.

c:\wp50\teach\glossary.epi
Updated 08/31/92