

Lay definition of common statistical / epidemiological terms

	Simplified formal definition	Lay definition
Screening		
Sensitivity	Proportion of true positives correctly identified as such by the test (Sensitivity = 1- False negative)	Proportion of individuals with disease that have positive test result e.g. sensitivity 66% \Rightarrow 2:3 of the individuals who have the disease will be identified by testing; 1:3 with disease will be testing negative and missed
False negative	True positives that are incorrectly identified as negative	Test result is negative in the presence of the disease
Specificity	Proportion of true negatives correctly identified as such by the test (Specificity= 1- False positive)	Proportion of individuals without disease that have negative test result e.g. specificity 90% \Rightarrow 9:10 of the individuals who don't have the disease will test negative; 1:10 without disease will be incorrectly labelled with disease
False positive	True negatives that are incorrectly identified as positive	Test result is positive in the absence of the disease
Positive predictive value	Proportion of identified positives that are truly positive Measure of process and feasibility of running a screening programme	Proportion of individuals with positive test who truly have the disease e.g. PPV 90% \Rightarrow 9:10 of those who tested positive have the disease; 1:10 test positive but don't have the disease
Adjustment/ standardisation		
Confounder	A variable that is associated both with the exposure and the outcome and it is not on the causal pathway between exposure and outcome. It gives an alternative explanation for an association between exposure and outcome. It is nuisance and need to be controlled	A factor (exposure) which can explain (entirely or partially) the observed association between 'exposure' and 'outcome' A confounder is a factor that distorts the association between 'exposure' and 'outcome'
Interaction	Effect modification It is the situation where the association between exposure and outcome varies according to	The association between 'exposure' and 'outcome' is of different strength in different strata of another factor e.g. the association between

	the level of a third factor It is real effect that needs to be detected and reported	smoking and lung cancer is stronger in older people than in younger people. In this association, age is the effect modifier
Crude estimate	A measure of effect which has not been altered to take into account the effect of confounding factors	Estimate derived without accounting for the effect of confounders (such as age, sex)
Adjusted estimate	A measure of effect which has been altered (by stratification or regression modelling) in order to take into account the effect of confounding factors	Controlled for the effect of confounders Taking into account 'factors' (such as age, sex, deprivation, etc) that distort the association between 'exposure' and 'outcome'
Regression analysis	Studies the association between two or more variables, where one is dependent on the other(s). This allows the dependent variable to be estimated given the value(s) of the other(s)	Statistical method that accounts simultaneously for multiple confounding factors
Standardisation	Direct (age) standardisation – observed age-specific rates of death or disease from the study population are applied to standard population of known age structure, thereby calculating age-adjusted rate Indirect (age) standardisation – age-specific death rates of standard population are applied to the age structure of the study population and compare total number of observed deaths with the number expected	A way of controlling for age (or other potential confounders) so that rates of disease or death in populations with different age structures (or other potential confounding factors) can be compared
DSR	Directly standardised rate is total number of deaths (that would have been expected if the study population had the same age structure as the standard population) divided by total person-time of the standard population	A method used to account for the difference in the age structure (or in other confounders) of populations in order to make a valid comparison of rates of disease/death

SMR	Standardised mortality ratio is the ratio of observed to expected deaths obtained by indirect method of standardisation (usually expressed as a %)	A method used to account for the difference in the age structure (or in other confounders) of populations in order to make valid comparison of death rates
Chance		
p-value	It is the result of hypothesis testing. It is the probability of obtaining the observed or more extreme estimate if the null hypothesis (H_0) were true. Small p-value suggests that the null hypothesis is unlikely to be true. Probability of making type I error i.e. wrongly rejecting H_0 in favour of H_1	The probability of having the observed estimate due to chance e.g. if $P < 0.05$ the difference observed occurs by chance alone less than 5 times in 100
95% Confidence interval	95% confident that the range of interval around the sample estimate contains the true population parameter If the sampling is repeated thousands of times, the interval around the sample estimate will include the true population parameter 95% of the time	95% confident that the true effect size lies within this range 'uncertainty range'
Power	Probability that an effect will be detected if it is truly there	Indicates how good a study is in identifying an effect (or difference in intervention) if in reality an effect (or difference) exists
Precision	Ability to measure magnitude of effect with minimal sampling error	Refers to the width of the confidence interval
Measures of frequency		
Prevalence	Number of existing cases in a given time / total population at risk	Proportion of existing cases

Incidence risk	Number of new cases during a period of time / population at risk at the start of the period	Proportion of new cases developed during a given time
Incidence rate	Number of new cases / total person-time at risk	Number of new cases over the sum of the different times each individual was at risk
Measures of effect		
Risk	It is the proportion of subjects who have experienced an outcome within a specified time period	Probability that an event will occur
Odds	Ratio of probability of occurrence of outcome to non-occurrence	Probability that an event will occur
Rate	Number of outcome events per unit person-time of follow up	Measure of effect (e.g. death) changing with time
Hazard	Rate defined at a precise point in time. Used in survival studies or Cox regression In survival studies, it is interpreted as the probability of dying at a particular point in time	Rate
Risk ratio	RR= Incidence risk in exposed / incidence risk in unexposed	Probability of having the 'outcome' among 'exposed' as compared to 'unexposed'
Odds ratio	In case control study: OR = odds of exposure among cases / odds of exposure among controls In cohort or intervention studies: OR = odds of outcome in exposed / odds of outcome in unexposed	e.g. OR of lung cancer with smoking of 1.3 \Rightarrow 30% increase in the likelihood of developing lung cancer among smokers as compared to non-smokers
Attributable fraction	Proportion of disease or outcome in the exposed individuals that can be attributed to the exposure (measure of effect) Attributable risk is used when there is causal association between exposure and outcome	Proportion of cases that would be avoided if the exposure is removed

Population attributable fraction	Proportion of disease or outcome in the entire population that can be attributable to the exposure (Measure of impact)	How much of the disease burden could be prevented by eliminating exposure
NNT	Number needed to treat	Number of patients that need to receive an intervention over a certain period to prevent one outcome
Measures of location		
Mean	Sum of the observations divided by the number of observations	Average
Mode	The most frequently occurring value in a set of data	Most common observation
Median	The value halfway through the dataset, below and above which lies an equal number of observation	Middle observation
Quintiles / Deciles	Measure of spread	Data broken down into 5 / 10 sections, each having 20% / 10% of the values. e.g. first quintile is the point with 20% of the data below it and 80% above it.
Outlier	An observation in a dataset that is substantially higher than lower than others in the dataset	A value that is outside the expected range
Measurement error		
Bias	Systematic error in the design or conduct of a study that results in conclusions that are different from truth It could be: Selection bias - error in selecting the study population such that comparison groups are not comparable Information bias - error in	Distorted truth due to error in selecting the study population or in measuring outcome/exposure

	measurement of exposure or outcome that results in systematic difference in the accuracy of the information collected between comparison groups. Includes reporter bias/ recall bias; interviewer bias / observer bias	
Validity	The extent to which a test measures the true value of the variable of interest	Test measuring what is intended to measure
Reliability	The extent to which a test will produce the same result if it is repeated at different periods or by different observers	Consistency of the test
Epidemiological terms		
Effectiveness	Effect of an intervention under operational conditions	Effect in real life condition
Efficacy	Effect of an intervention under trial conditions	Effect in ideal conditions
Efficiency	Effective outcome with use of minimal resources in terms of time and money	Maximum output with minimum cost
Systematic review	Systematic assembly, critical appraisal and synthesis of all relevant studies on particular topic	Systematically identifying, selecting and critically appraising of literature on topic of interest
Meta-analysis	A statistical method of pooling information from systematically selected studies and quantitative synthesis of summary result. It is used in order to give an accurate estimate of effect and to overcome problems of reduced statistical power of small sample size studies	A statistical method used to combine information from studies addressing the same research question to give an overall summary estimate
Blinding	Concealment of information about exposure or outcome in order to reduce bias	Participants and/ or investigators don't know to which intervention they have been allocated
Randomisation	A procedure in which the probability of allocation to a	Random allocation of participants of a study to the intervention or the

	<p>particular group is predetermined and each individual has equal chance of being allocated to all groups. Randomisation ensures that intervention and control groups are similar with respect to known and unknown confounders, and it prevents bias in the allocation of participants to intervention or control groups</p>	control group
Intention to treat analysis	<p>Used in intervention studies, comparing intervention and control groups as they were originally randomised, irrespective of whether some people in each group stopped taking the allocated treatment, changed to other treatment, or were lost to follow up</p>	<p>Analysis according to the groups that participants were allocated to, whether or not they actually received the intervention</p>

Note: The lay term explanations would be clearer if used in context rather than in abstract as presented above