Absolute measures of association	Measures of association that estimate the impact of an exposure on the population. They include measures such as the risk difference and the incidence rate difference.
Allocation concealment	In Randomised Controlled Trials, the treatment is allocated randomly to each participant. Allocation concealment means tha none of the people directly involved in the study knows which patient will receive which treatment, as knowledge of the next assignment could influence whether a patient is included or excluded based on perceived prognosis.
Alternative hypothesis	The alternative hypothesis is the converse of the null hypothesis. The alternative hypothesis is often that a difference between groups does exist. If the null is rejected due to a small p-value, then we can accept the alternative. If the null hypothesis is not rejected using statistical inference, we cannot assume that the alternative hypothesis holds. Instead, we can only conclude there was not enough evidence to reject the null hypothesis.
Antagonism	When there is effect modification in a study, antagonism means that the effect modifier diminishes the effect of the exposure on the outcome.
Attributable risk	A measure of association that indicates, on an absolute scale, how much greater the frequency of disease in the exposed group is compared with the unexposed, assuming the relationship between exposure and outcome is causal (an important assumption). It is the difference between the incidence rate or the risk in the exposed and non exposed groups.
Attributable risk percent	A measure that indicates what proportion of the risk (or incidence rate) in the exposed group can be attributed to the exposure of interest, assuming the relationship between exposure and outcome is causal.
Bias	Bias is any trend in the collection, analysis, interpretation, publication or review of data that can lead to conclusions that are systematically different from the truth.
Blinding	A procedure whereby one or more parties in a trial are kept unaware of which treatment arm participants have been assigned to. Single blinding occurs when only one party is blinded in a trial, usually the participants. Double blinding occurs when both the participants and study staff are blinded in a trial.
Case	A case is an individual with the outcome under study. Epidemiological research is based on the ability to quantify the occurrence of disease in populations. This requires a clear definition of what is meant by a case. This could be a person who has the disease, health disorder, or suffers the event of interest. By “event” we mean a change in health status, (e.g. death in studies of mortality or becoming pregnant in fertility studies).
Case-cohort study	Case-cohort is a nested study design, therefore requires an existing cohort study. In a case-cohort study, cases are defined as those participants of the cohort who developed the disease of interest, but the control group is selected from all cohort participants at baseline before the cases develop.
Case-control study	Study in which individuals are selected on the basis of whether or not they have the outcome of interest; usually some relatively rare outcome. Exposure (risk factor) status is explored to establish whether the exposure is more common in the case (those that have the outcome) or control (those that do not have the outcome) group. This type of study always results in an odds ratio, for example comparing the odds of being exposed (e.g. a smoker) in those who had the outcome (e.g. pancreatic cancer), with the odds of being a smoker in those who did not have pancreatic cancer.
Causation	The key question in most medical research. Research works by trying to disprove alternative explanations (e.g. chance, confounding). If this can be done, then the relationship between the exposure and the outcome will be one of causation.
Chi-squared test	This is a statistical procedure for testing whether two proportions are similar.
Cohort study	A cohort study typically involves group of people without disease, exposed and unexposed to the exposure of interest, who are observed over a period of time to see what happens to them and is also known as a longitudinal study.
Confidence interval (95%)	This is an estimated range of values calculated from a given set of sample data which are likely to contain the ‘true’ population value. The range of values around a relative risk measure which would, in 95% of such studies, contain the ‘true’ risk (the true risk being the relative risk that would be obtained if the study had included the entire population of patients). By “contain the true value”, we mean that the true value lies above the lower value of the confidence interval but below the upper values of the confidence interval.
Confounding	Confounding is the effect of an extraneous variable that wholly or partially accounts for the apparent effect of the study exposure or that masks an underlying true association.
Control (as opposed to a case)	A control is a person without the outcome under study (in a type of epidemiological study called a case-control study) or a person not receiving the intervention (in a clinical trial). The choice of an appropriate group of controls requires care, as we need to be able to draw useful comparisons between these controls and the cases/intervention group.
Cumulative incidence	Cumulative incidence (or risk) measures the occurrence of disease in the population and is defined as the proportion of the population with a new event during a given time period. To calculate cumulative incidence, we divide the number of new cases during the period of interest by the number of disease-free individuals at the start of this time period.
Difference	Differences are absolute measures of association. They include risk difference and incidence rate difference
Directed Acyclic Graphs	The Directed Acyclic Graphs present causal associations between variables and can be used to identify sources of bias and confounding in a study.

Effect modification	Effect modification exists when the strength of the association between an exposure and an outcome varies over different levels of a third variable.
Exposure	When people have been ‘exposed’, they have been in contact with something that is hypothesised to have an effect on health, which can be either positive or negative e.g. tobacco, nuclear radiation, pesticides in food (all negative effects), physical exercise and eating fruit and vegetables (all positive effects). This is the most obvious meaning of 'exposed', but it can also refer to any patient characteristic or risk factor for the outcome of interest.
External validity	A study has external validity (or generalisability) when inferences based on the study findings are applicable to individuals outside the source population.
Generalisibility	Another term for external validity.
Gold standard	A recognised way of determining who really has the disease.
Hypothesis	A statement that can be tested using quantitative evidence (data) in a hypothesis test, the foundation of modern science.
Incidence rate	Incidence rate a measure of frequency and is defined as the number of new cases per unit of person-time. To calculate it, you need to divide the number of new cases recorded during the follow-up period by the total person-time contributed by disease-free individuals during the same period.
Incidence rate ratio	A relative measure of association. To calculate it, you divide the incidence rate in the exposed group by the incidence rate in the unexposed group.
Information bias	Information bias is the type of bias that is the result of misclassification of the exposure or the disease status or both.
Intention to treat	In RCTs, analysing patients according to which group they were originally assigned, regardless of whether they followed the allocated treatment or not, is called intention-to-treat. This ensures that randomisation is preserved during the analysis.
Interaction	An interaction occurs when a predictor variable has a different effect on the outcome depending on the value of another predictor variable. This is also called effect modification in epidemiology.
Internal validity	A study has internal validity when the study findings truly reflect the association between the exposure and the outcome in the source population.
Lead time bias	This is a type of bias that may occur when evaluating screening programs. Individuals who were diagnosed through screening and those who were diagnosed because they experienced symptoms may have equal survival, but screened cases could seem to have survived longer, because they were diagnosed earlier.
Length time bias	This is a type of bias that may occur when evaluating screening programs. Screening may preferentially identify slower-growing or less progressive cases of a disease that have better prognosis. These cases would have survived longer anyway, so when you compare screened and not screened cases, the seemingly improved survival of those diagnosed through screening may not reflect the truth.
Loss to follow up	Study participants are often followed up through periodic assessment (questionnaires, tests etc.). Some participants may relocate, die, withdraw their consent to participate or become untraceable for other reasons. This phenomenon is known as loss to follow up and may introduce bias in the study.
Matching	A method for “controlling for” (e.g. effectively removing) the effect of confounding at the design stage of a case-control study; controls are selected to have a similar distribution of potentially confounding variables to the cases, e.g. they are said to be “matched” for sex if there are similar proportions of men and women in both groups.
Measures of association	Measures that quantify the association between an exposure and an outcome. They can be either absolute or relative measures.
Misclassification	Misclassification means that certain individuals are wrongly assigned to a certain group. For example, individuals exposed to the exposure of interest may be wrongly assigned to the unexposed group or individuals with the disease may be wrongly assigned to the non-diseased group. Misclassification may be differential or non-differential, with different implications for the validity of the study.
Negative predictive value	The proportion of negative test results that are in people who do not have the disease
Nested case-control study	Nested case-control is a nested study design, therefore requires an existing cohort study. In a nested case-control study, cases are defined as those participants of the cohort who developed the disease of interest, but the control group is selected among the cohort study participants.
Null hypothesis	The null hypothesis is what the investigator sets out to disprove in order to find evidence of an association between two or more things. The null hypothesis is often that there is no difference between patient groups regarding the outcome of interest. The null hypothesis can then be rejected using statistical tests and their associated p-values
Number needed to treat	Number needed to treat is the inverse of the attributable risk and it's an expression of the number of people that need to receive the intervention in order to prevent one case of the disease.
Odds	The odds is a way to express probability, e.g. the odds of exposure is the number of people who have been exposed (or have a disease) divided by the number of people who have not been exposed (or do not have the disease). The mathematical relationship between odds and probability is: Odds = probability / (1 – probability)

Odds ratio	The odds ratio is a relative measure of association. It can be calculated by dividing the odds of an exposure or a disease in one group by the odds of the exposure or the diseases in the comparison group.
Outcome	The event or main quantity of interest in a particular study, e.g. death, contracting a disease, blood pressure.
Person-time	Person-time is a measure of the time spent in the study by the participants. Participants contribute person-time from the moment they enter the study until they get the outcome of interest, they die or they are lost to follow-up. Person-time is used in the calculation of the incidence rate.
Population	The set of all people of interest to a study. We can't study them directly and so must instead draw a sample of people from the population.
Population attributable risk	A measure of association that indicates, on an absolute scale, how much greater the frequency of disease in the population is compared with the unexposed, assuming the relationship between exposure and outcome is causal (an important assumption). It is the difference between the incidence rate or the risk in the population and the unexposed group.
Population attributable risk percent	A measure that indicates what proportion of the risk (or incidence rate) in the population can be attributed to the exposure of interest, assuming the relationship between exposure and outcome is causal.
Positive predictive value	The proportion of positive test results that occur in people who actually have the disease.
Pre-clinical phase	Pre-clinical phase of a disease is the period, before symptoms appear, that the disease could be detected; for instance, an indicator in the blood might be elevated. This period is important for screening.
Prevalence	Prevalence is the proportion of individuals in a population who have the disease (or attribute of interest) at a specific time point. To calculate prevalence, we divide the number of people with the disease by the total number of individuals in the population.
Prevention paradox	A large number of people exposed to low risk generate more cases than a small number of people exposed to high risk. Therefore, a measure that brings large benefits to the community offers little to each participating individual. This is called the prevention paradox.
p-value	This is the probability of obtaining the study result (relative risk, odds ratio etc.) or one that's more extreme - if the null hypothesis is true. The smaller the p-value, the easier it is for us to reject the null hypothesis and accept that the result was not just due to chance. A p-value of <0.05 means that there is only a very small chance of obtaining the study result if the null hypothesis is true, and so we would usually reject the null. Such as result is commonly called “statistically significant”. A p-value of >0.05 is usually seen as providing insufficient evidence against the null hypothesis, so we accept the null.
Randomisation	A method for ensuring that both groups in a clinical trial (e.g. those receiving the intervention and those not receiving the intervention (controls), have similar proportions of confounding variables, such as age.
Rate	A rate expresses how quickly the outcome of interest occurs, so is subtly different from a risk (even if many non-epidemiologists use the two words interchangeably). The denominator is some measure of person-time
Ratio	The term ratio describes a number obtained by dividing one quantity by another. Relative measures of association are ratios.
Recall bias	Recall bias is the bias caused by the differentially inaccurate recall of past exposure between cases and controls.
Relative measures of association	Measures of association that estimate the strength pf the association. They include measures such as the risk ratio, the incidence rate ratio and the odds ratio.
Restriction	A method for controlling for confounding at the design stage of a study, e.g. by including patients in a clinical trial only between the ages of 45 and 65 without so that the results of the trial are not confounded by different levels of age in the two treatment groups.
Risk	Risk (or cumulative incidence measures the occurrence of disease in the population and is defined as the proportion of the population with a new event during a given time period. To calculate risk, we divide the number of new cases during the period of interest by the number of disease-free individuals at the start of this time period.
Risk ratio	The risk ratio is a relative measure of association. It can be calculated by dividing the risk of the disease in the exposed group by the risk of the disease in the unexposed group.
Risk difference	A measure of association that indicates, on an absolute scale, how much greater the frequency of disease in the exposed group is compared with the unexposed. It is the difference between the risk in the exposed and non exposed groups.
Risk factor	Something that increases a person’s chance of getting a disease.
Sample	A sample is a relatively small number of observations (or patients) from which we try to describe the whole population from which the sample has been taken. Typically, we calculate the mean for the sample and use the confidence interval to describe the range within which we think the population mean lies. This is one of the absolutely key concepts behind all medical research (and much non-medical research too).
Sample size	(Usually) The number of people in your sample.
Selection bias	Selection bias is the type of bias that occurs when an individual’s probability of being included in the study sample is related to both the exposure and the outcome.
Sensitivity	Sensitivity is defined as the proportion of those with the disease who tested positive. In other words, the proportion of the true positive results among all individuals with the disease.

Specificity	Specificity is defined as the proportion of those without the disease who tested negative. Alternatively, the proportion of true negative results among all those without the disease.
Statistical test	This is the only way to decide whether the results of your analysis (e.g. your measure for group A compared with your measure for group B, are likely to be due to chance or could be real).
Stratification	A method for controlling for confounding at the analysis stage of a study. Risks are calculated separately for each category of confounding variable (e.g. each age group and each sex separately) and then a weighted pooled estimate is calculated.
Synergism	When there is effect modification in a study, synergism means that the effect modifier potentiates the effect of the exposure on the outcome.
Two-by-two table	A 2x2 table is an easy way to represent the groups of exposed/unexposed and diseased/non-diseased participants in a study. A 2x2 table provides a clear format to present the data and makes calculation of measures of frequency and association much simpler
Variable	A variable is a characteristic or item that can take different values. They can be categorical or numerical variables: for example, disease stage or age.
Synergism	When there is effect modification in a study, synergism means that the effect modifier potentiates the effect of the exposure on the outcome.
Two-by-two table	A 2x2 table is an easy way to represent the groups of exposed/unexposed and diseased/non-diseased participants in a study. A 2x2 table provides a clear format to present the data and makes calculation of measures of frequency and association much simpler
Variable	A variable is a characteristic or item that can take different values. They can be categorical or numerical variables: for example, disease stage or age.