Research Terms Glossary

Professor Khalid Khan

Distinguished Investigator, Editor, and Author of award-winning book *Systematic reviews to support Evidence-based Medicine*

Khalid Khan's Glossary of Research Terms

Absolute risk reduction (ARR) see Risk difference (RD).

Accuracy measure A statistic for summarising the accuracy with which a test predicts a diagnosis. There are three commonly used sets of accuracy measures for binary tests: sensitivity and specificity; positive and negative predictive values; and likelihood ratios. All these measures are paired. Single measures of accuracy are seldom used with the exception of the diagnostic odds ratio.

Adverse effect It is an undesirable and unintended harmful or unpleasant reaction resulting from an intervention. It often predicts hazard from its future administration and warrants prevention or specific treatment or alteration or withdrawal of the intervention.

Attrition bias (exclusion bias) Systematic differences between study groups caused by exclusion or dropout of subjects (e.g. because of side effects of intervention) from the study. Intention-to-treat analysis in combination with appropriate sensitivity analyses including all subjects can protect against this bias. Also see Intention-to-treat (ITT) analysis and Withdrawals.

Baseline risk The frequency of outcome in a population without intervention. It is related to the severity of underlying disease and prognostic features. Good prognosis is associated with low baseline risk while poor prognosis is associated with high baseline risk of undesirable outcomes. Baseline risk is important for determining who will benefit most from interventions. Also see Number needed to treat (NNT).

Bias (systematic error) A tendency for results to depart systematically, either lower or higher, from the 'true' results. Bias either exaggerates or underestimates the 'true' effect of an intervention or exposure. It may arise due to several reasons, e.g. errors in design and conduct of a study. This may lead to systematic differences in comparison groups (selection bias), differences in care or exposure to factors other than the intervention of interest (performance bias), differences in assessment of outcomes (measurement bias), withdrawals or exclusions of people entered into the study (attrition bias), etc. Studies with unbiased results are said to be internally valid.

Binary data Measurement where the data have one of two alternatives, for example the patient is either alive or dead, the test result is either positive or negative, etc.

Bivariate model A statistical method for generating summary estimates of test accuracy. It adjusts for the correlation that might exist between test sensitivity and specificity.

Blinding (masking) Blinding keeps the study participants, caregivers, researchers and outcome assessors ignorant about the interventions to which the subjects have been allocated in a study. In single blind studies only the subjects are ignorant about interventions, whilst in double blind studies both the participants and caregivers or researchers are blind. Outcome assessors can often be blinded even when participants and caregivers can't be. Blinding protects against performance bias and detection bias, and it may contribute to adequate allocation concealment during randomisation. Also see Randomisation.

Boolean logic Boolean logic (named after George Boole) refers to the logical relationship among search terms. Boolean operators AND, OR and NOT are used during literature searches to include or exclude certain citations from electronic databases. They are also used in Internet search engines.

Case-control study A comparative observational study where participants/patients with the outcome (cases) and those without the outcome (controls) are compared for their prior intervention or exposure rates.

Clinical trial A loosely defined term generally meaning to describe a study to evaluate efficacy and effectiveness of interventions. This term encompasses study designs ranging from randomised controlled trials to uncontrolled observations of a few cases.

Cochrane Collaboration An international not-for-profit organisation that aims to help with informed decision-making about health care by preparing, maintaining and improving accessibility of systematic reviews of interventions (http://www.cochrane.org). The major product of the Collaboration is the Cochrane Database of Systematic Reviews which is part of the Cochrane Library (http://www.update-software.com/cochrane/). Those who prepare Cochrane Reviews are mostly health care professionals who volunteer to work in one of more than forty Collaborative Review Groups (CRGs). Each CRG has a coordinator, and an editorial team to oversee the quality of their reviews. The activities of the Collaboration are directed by an elected Steering Group and are supported by staff in Cochrane Centres worldwide.

Cohort study A comparative observational study where participants with an intervention or exposure (not allocated by the researcher) are followed up to examine the difference in outcomes compared to a control group, e.g. those receiving no care.

Comparative study A study where the effect of an intervention or exposure is assessed using comparison groups. This can be a randomised controlled trial, a cohort study, a case-control study, etc.

Confidence interval (CI) The range within which the "true" value of a measurement (e.g. effect of an intervention) is expected to lie in a population with a given degree of certainty. Confidence intervals represent the distribution probability of random errors, but not of systematic errors (bias). Conventionally, 95% confidence intervals are used.

Confounding A situation in studies where the effect of an intervention on an outcome is distorted due to the association of the outcome with another factor, the confounding variable, which can prevent or cause the outcome independent of the intervention. It occurs when groups being compared are different with respect to important factors other than the interventions or exposures under investigation. Adjustment for confounding requires stratified or multivariable analysis. Also see Randomisation.

Continuous data Measurement on a continuous scale such as height, weight, blood pressure, etc. For continuous data, effect is often expressed in terms of mean difference. Also see Effect size (ES).

Control event rate (CER) The proportion of subjects in the control group in whom an event or outcome is observed, in a defined time period.

Controlled clinical trial A loosely defined term to describe a prospective comparative study for assessing effectiveness of interventions (regardless of whether randomisation is used or not). Watch out for indiscriminate use of this ambiguous term in reviews. It is also a MeSH in the Medline database.

Cost-effectiveness analysis see Economic evaluation and Efficiency.

Diagnostic odds ratio The ratio of the likelihood ratio for a positive test result to the likelihood ratio for a negative test result. It provides a single measure of accuracy. Also see Accuracy measure.

Diary keeping A qualitative research method, usually an addition to questionnaire or interview data, where participants record experience and emotions contemporaneously. It can be free form, where people write what they want to, or structured where they have specific questions to answer or topics to write about.

Dose-response A dose-response relationship demonstrates that at higher doses the strength of association between exposure and outcome is increased.

Economic evaluation (e.g. cost-effectiveness analysis) A study that takes into account both the clinical effectiveness and the costs of alternative interventions to address the question of how to achieve an optimal clinical outcome at least cost. The term cost-effectiveness analysis is often used synonymously but this is a misnomer. A full economic evaluation considers both clinical and cost outcomes whereas a partial evaluation may only consider costs without regard to clinical outcomes.

Effect (effect measure, treatment effect, estimate of effect, effect size) Effect is the observed association between interventions and outcomes or a statistic to summarise the strength of the observed association. The statistic could be a relative risk, odds ratio, risk difference, or number needed to treat for binary data; a mean difference, or standardised mean difference for continuous data; or a hazard ratio for survival data. The effect has a point estimate and a confidence interval. The term *individual effect* is often used to describe effects observed in individual studies included in a review. The term *summary effect* is used to describe the effect generated by pooling individual effects in a meta-analysis.

Effect modification It occurs when a factor influences the effect of the intervention under study e.g. age may modify responsiveness to treatment.

Effect size (ES) This term is sometimes used for an effect measure for continuous data. Also see Effect measure.

Effectiveness The extent to which an intervention (therapy, prevention, diagnosis, screening, education, social care, etc.) produces a beneficial outcome in the routine setting. Unlike efficacy, it seeks to address the question: Does an intervention work under ordinary day-to-day circumstances?

Efficacy The extent to which an intervention can produce a beneficial outcome under ideal circumstances.

Efficiency The extent to which the balance between input (costs) and outputs (outcomes) of interventions represents value for money. It addresses the question of whether clinical outcomes are maximised for the given input costs. Also see Economic evaluation.

Evidence-based medicine (EBM) The conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. It involves the process of systematically finding, appraising, and using contemporaneous research findings as the basis for clinical decisions. Evidence-based practice (EBP) is a related term. Both EBM and EBP follow four steps: formulate a clear clinical question from a patient's problem; search the literature for relevant clinical articles; evaluate (critically appraise) the evidence for its validity and usefulness; implement useful findings in clinical practice taking account of patients' preferences and caregivers' experience. Another related term is evidence-based health care, which is an extension of the principles of EBM to all professions associated with health care, including purchasing and management. Systematic reviews provide powerful evidence to support all forms of EBM.

Experimental event rate (EER) The proportion of participants in the experimental group in whom an event or outcome is observed, in a specified time period.

Experimental study A comparative study in which decisions concerning the allocation of participants or patients to different interventions are under the control of the researcher e.g. randomised controlled trial.

Exposure A factor (including interventions) which is thought to be associated with the development or prevention of an outcome.

External validity (generalisability, applicability) The extent to which the effects observed in a study can be expected to apply in routine clinical practice, i.e. to people who did not participate in the study. Also see Validity.

Fixed effect model A statistical model for combining results of individual studies, which assumes that the effect is truly constant in all the populations studied. Thus, only within-study variation is taken to influence the uncertainty of the summary effect and it produces narrower confidence intervals than the random effects model. Also see Random effects model.

Focus group A qualitative research method. The collection of qualitative data using a group interview on a topic. Usually 6 to 12 participants are involved, and they can be used to gauge issues of importance. Also see Interview.

Forest plot A graphical display of individual effects observed in studies included in a systematic review along with the summary effect, if meta-analysis is used.

Funnel plot A scatter plot of effects observed in individual studies included in a systematic review against some measure of study information, e.g. study size, inverse of variance, etc. It is used in exploration for the risk of publication and related biases.

Generalisation The extent to which findings of a qualitative research study are consistent with findings of similar studies, adding to the understanding of a phenomenon. Within qualitative research the aim is not to extrapolate to wider populations, so this term should not be confused with the terms External validity and Generalisability.

Generalisability See External validity. Also see Generalisation.

GRADE The Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group is an informal collaboration that aims to develop a comprehensive methodology for assessing the strength of the evidence collated in systematic reviews and for generating recommendations from evidence in guidelines. See www.gradeworkinggroup.org.

Guidelines Statements that aim to assist practitioners and patients in making decisions about specific clinical situations. They often, but not always, use evidence from systematic reviews.

Hazard ratio An effect measure for survival data, which compares the survival experience of two groups.

Health technology assessment (HTA) Health technology includes any method used by those working in health services to promote health, to screen, diagnose, prevent and treat disease, and to improve rehabilitation and long-term care. HTA considers the effectiveness, appropriateness, costs and broader impact of interventions using both primary research and systematic reviews.

Heterogeneity/Homogeneity The degree to which the effects among individual studies being systematically reviewed are similar (homogeneity) or different (heterogeneity). This may be observed graphically by examining the variation in individual effects (both point estimates and confidence intervals) in a forest plot. Quantitatively, statistical tests of heterogeneity/homogeneity may be used to determine ifdsf the observed variation in effects is greater than that expected due to the play of chance alone. For making a clinical judgement about heterogeneity, one might look at the differences between populations, interventions and outcomes of studies.

Homogeneity see Heterogeneity.

I_2 statistic It is a statistic for assessment of heterogeneity during study synthesis. Ranging from 0% to 100%, it gives the percentage of total variation across studies due to heterogeneity.

Intention-to-treat (ITT) analysis An analysis where subjects are analysed according to their initial group allocation, independent of whether they dropped out or not, fully complied with the intervention or not, or crossed over and received alternative interventions. A true ITT analysis includes an outcome (whether observed or estimated) for all patients. Also see Attrition bias and Sensitivity analysis.

Internal validity see Validity.

Intervention A therapeutic or preventative regimen e.g. a drug, an operative procedure, a dietary supplement, an educational leaflet, a test (followed by a treatment), etc. undertaken with the aim of improving health outcomes. In a randomised trial, the effect of an intervention is the comparison of outcomes between two groups, one with the intervention and the other without (e.g. a placebo or another control intervention).

Interview A qualitative research method. It involves questioning people about their views or experience of a phenomenon or event. Can range from structured, where the same questions are asked to each participant to unstructured, which consists of a list of broad areas to be covered, the exact format of each interview being determined as it progresses. Also see focus group.

Inverse of variance see Variance.

Kirkpatrick hierarchy A classification of medical educational outcomes to capture of impact of educational interventions. It has various levels: 1a: Participation or completion captures attendance at and views on the learning experience, e.g. course evaluation; 1b: Modification of attitudes captures change in attitudes or perceptions, e.g. subjective reaction or satisfaction of participants with course, difference between pre and post course attitude questionnaire; 2: Modification of knowledge or skills captures change knowledge or skills, e.g. difference in scores from pre to post course; 3: Health professional's behaviour captures the transfer of learning to the workplace or integration of new knowledge and skills leading to modification of behaviour or performance, e.g. difference in performance after the teaching evidenced by more evidence based prescribing and more frequent attendance at journal club; and 4: Change in delivery of care and health outcomes captures changes in the delivery of care attributable to the educational programme with or without assessments of improvement in the health outcomes and wellbeing of patients as a direct result of teaching, e.g. audit of practice showing greater compliance with evidence based criteria

Likelihood ratio (LR) It is the ratio of the probability of a positive (or negative) test result in subjects with disease to the probability of the same test result in subjects without disease. The LR indicates by how much a given test result raises or lowers the probability of having the disease. With a positive test result, a LR+ >1 increases the probability that disease will be present. The greater the LR+, the larger the increase in probability of the disease and the more clinically useful the test result. With a negative test result, a LR- <1 decreases the probability that the disease is present: the smaller the LR-, the larger the decrease in the probability of disease and the more clinically useful the test result.

Mean difference The difference between the means (i.e. the average values) of two groups of measurements on a continuous scale. Also see Effect, Standardised mean difference (SMD).

Measurement bias (detection bias, ascertainment bias) Systematic differences between groups in how outcomes are assessed in a study. Blinding of study subjects and outcome assessors protects against this bias.

MeSH Medical Subject Heading. Controlled terms used in the MEDLINE database to index citations. Other electronic bibliographic databases frequently use MeSH-like terms.

Meta-analysis A statistical technique for combining (pooling) the results of a number of studies addressing the same question to produce a summary result.

Meta-regression A multivariable model with effect estimates of individual studies (usually weighted according to their size) as dependent variable and various study characteristics as independent variables. It searches for the influence of study characteristics on the size of effects observed in a systematic review. Also see Multivariable analysis.

Metasynthesis The amalgamation of the results of a group of qualitative studies on the same or a related issue. Included studies can be evaluated and the findings combined. This is achieved from reviewing the published data and not from meta-analysing data.

Multivariable analysis (multivariable model) An analysis that relates some independent or explanatory or predictor variables (X_1, X_2, ...) to an dependent or outcome variable (Y) through a mathematical model such as $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots$, where Y is the outcome variable Y; β_0 is the intercept term; and β_1, β_2, ... are the regression coefficients indicating the impact of the independent variables X_1, X_2, ... on the dependent variable Y. The coefficient is interpreted as the change in the outcome variable associated with a one-unit change in the independent variable and provides a measure of association or effect. Multivariable analysis is used to adjust for confounding e.g. by including confounding factors along with the intervention (or exposure) as the independent variables in the model. This way the effect of intervention (or exposure) on outcome can be estimated while adjusting for the confounding effect of other factors. Also see Confounding.

Negative predictive value The proportion of subjects who test negative who truly do not have the disease.

Normal distribution A frequency distribution that is symmetrical around the mean and bell shaped (also called Gaussian distribution).

Null Hypothesis The hypothesis put forward when carrying out significance tests that states that there is no difference between groups in a study. For example, statistically we discover that an intervention is effective by the rejecting the null hypothesis that outcomes are not different between the experimental and the control group. Also see p-value.

Number needed to harm (NNH) It is the number of patients treated for whom there is one additional patient who experiences an episode of harm (adverse effect, complication, etc.). It is computed in the same manner as NNT.

Number needed to treat (NNT) An effect measure for binary data. It is the number of patients who need to be treated to prevent one undesirable outcome. In an individual study it is the inverse of risk difference (RD). In a systematic review it is computed using baseline risk and a measure of relative effect (relative risk, odds ratio). It is a clinically intuitive measure of the impact of a treatment.

Observational study Research studies in which interventions, exposures and outcomes are merely observed with or without control groups. These could be cohort studies, case-control studies, cross-sectional studies, etc.

Odds The ratio of the number of participants with an outcome to the number without the outcome in a group. Thus, if out of 100 subjects, 30 had the outcome (and 70 did not), and the odds would be 30/70 or 0.42. Also see Risk.

Odds ratio (OR) An effect measure for binary data. It is the ratio of odds of an event or outcome in the experimental group to the odds of an outcome in the control group. An OR of one indicates no difference between comparison groups. For undesirable outcomes an OR that is less than one indicates that the intervention is effective in reducing the odds of that outcome. Also see Relative risk.

Outcome The changes in health status that arise from interventions or exposure. The results of such changes are used to estimate the effect.

p - value (statistical significance) The probability, give a null hypothesis, that the observed effects or more extreme effects in a study could have occurred due to play of chance (random error). In an effectiveness study, it is the probability of finding an effect by chance as unusual as, or more unusual than, the one calculated, given that the null hypothesis is correct. Conventionally, a p-value of less than 5% (i.e. $p<0.05$) has been regarded as statistically significant. This threshold, however, should never be allowed to become a straight jacket. When statistical tests have low power, e.g. tests for heterogeneity, a less stringent threshold (e.g. $p<0.1$ or <0.2) may be used. Conversely, when there is a risk of spurious significance, e.g. multiple testing in subgroup analysis, a more stringent threshold (e.g. $p<0.01$) may be used. When interpreting the significance of effects, p-values should always be used in conjunction with confidence intervals (CI). Also see Confidence interval.

Performance bias Systematic differences in the care provided to the study subjects apart from the interventions being evaluated. Blinding of carers and subjects and standardisation of the care plan can protect against this bias.

Phenomenon An occurrence or a fact. Phenomenon is often used as a generic term for the object of a qualitative research study.

Point estimate of effect The observed value of the effect of an intervention among the subjects in a study sample. Also see Confidence interval (CI).

Positive predictive value The proportion of subjects who test positive who truly have the disease.

Posttest probability of disease An estimate of the probability of disease in light of the information obtained from testing. With accurate tests, the post-test estimates of probabilities change substantially from pre-test estimates. In this way a positive test result may help to rule in disease and a negative test result may help to rule out disease.

Power The ability to demonstrate an association when one exists. The ability to reject the null hypothesis when it is indeed false. Power is related to sample size. The larger the sample size, the more the power, and the lower is the risk that a possible association could be missed.

Precision (specificity) of a search The proportion of relevant studies identified by a search strategy expressed as a percentage of all studies (relevant and irrelevant) identified by that method. It describes the ability of a search to exclude irrelevant studies. Also see Sensitivity of a search.

Precision of effect see Random error.

Pretest probability of disease An estimate of probability of disease before tests are carried out. It is usually estimated as the prevalence of disease in a given setting (e.g. community, primary care, secondary care, hospital, etc.) Sometimes, when such information is not available, it may have to be estimated.

Publication bias Arises when the likelihood of publication of a study is related to the significance of its results. For example, a study is less likely to be published if it finds an intervention ineffective. Reviewers should make all efforts to identify such negative studies otherwise their inferences about the value of intervention will be biased. Funnel plots may be used to explore for the risk of publication and related biases.

Qualitative research Research concerned with the subjective world that offers insight into social, emotional and experiential phenomena in health and social care. Including findings from qualitative research may enhance the quality and salience of reviews.

Quality of a qualitative research study The quality of a qualitative research study depends on the degree to which its design, conduct and analysis is trustworthy. Trustworthiness consists of several concepts including credibility, dependability, transferability, confirmability.

Quality of a study (methodological quality) The degree to which a study minimises biases. Features related to the design, the conduct and the statistical analysis of the study can be used to measure quality. This determines the validity of results.

Quasi-experimental (quasi-randomised) study A term sometimes used to describe a study where allocation of subjects to different groups is controlled by the researcher, like in an experimental study, but the method falls short of genuine randomisation (and allocation concealment), e.g. by using date of birth or even-odd days.

Random effects model A statistical model for combining the results of studies that allows for variation in the effect among the populations studied. Thus, both within-study variation and between-studies variation are included in the assessment of the uncertainty of results. Also see Fixed effect model.

Random error (imprecision or sampling error) Error due to the play of chance that leads to wide confidence intervals around point estimates of effect. The width of the confidence interval reflects the magnitude of random error or imprecision. Also see p-value.

Randomisation (with allocation concealment) Randomisation is the allocation of study subjects to two or more alternative groups using a chance procedure, such as computer generated random numbers, to generate a sequence for allocation. It ensures that subjects have a prespecified (very often an equal) chance of being allocated one of two or more interventions. In this way the groups are likely to be balanced for known as well as unknown and unmeasured confounding variables. Concealment of the allocation sequence until the time of allocation to groups is essential for protection against selection bias. Foreknowledge of group allocation leaves the decision to recruit the subject open to manipulation by researchers and study subjects themselves. Allocation concealment is almost always possible even when blinding is not. Randomisation alone without concealment does not protect against selection bias.

Randomised controlled trial (RCT) A comparative study with random allocation (with allocation concealment) of subjects to interventions, and follow up to examine differences in outcomes between the various groups.

Relative Risk (RR) (risk ratio, rate ratio) An effect measure for binary data. It is the ratio of risk in the experimental group to the risk in the control group. An RR of one indicates no difference between comparison groups. For undesirable outcomes an RR that is less than one indicates that the intervention is effective in reducing the risk of that outcome. Also see Odds ratio.

Review An article that summarises the evidence contained in a number of different individual studies and draws conclusions about their findings. It may or may not be systematic. Also see Systematic Review and Meta-analysis.

RevMan The Cochrane Collaboration's software for review management and meta-analysis available at http://www.cochrane.org/cochrane/revman.htm.

Risk (proportion or rate) The proportion of subjects in a group who are observed to have an outcome. Thus, if out of 100 subjects, 30 had the outcome, the risk (rate of outcome) would be 30/100 or 0.30. Also see Odds.

Risk difference (RD) (absolute risk reduction, rate difference) An effect measure for binary data. In a comparative study, it is the difference in event rates between two groups. The inverse of RD produces number needed to treat (NNT). Also see Number needed to treat.

Sample Subjects selected for a study from a much larger group or population.

Selection bias (allocation bias) Systematic differences in prognosis and/or therapeutic sensitivity at baseline between study groups. Randomisation (with concealed allocation) of a large number of patients protects against this bias.

Sensitivity (recall) of a search The proportion of relevant studies identified by a search strategy expressed as a percentage of all relevant studies on a given topic. It describes the comprehensiveness of a search method, i.e. its ability to identify all relevant studies on a given topic. Highly sensitive strategies tend to have low levels of specificity (precision) and vice versa. Also see Precision of a search.

Sensitivity (true positive rate) of a test The proportion of those people who really have the disease who are correctly identified as such.

Sensitivity analysis Repetition of an analysis under different assumptions to examine the impact of these assumptions on the results. In systematic reviews, when there is poor reporting in individual studies authors of primary studies should be asked to provide missing and unclear information. However, this is not always possible and reviewers often have to make assumptions about methods and data and they may impute missing information. In this situation, sensitivity analysis should be carried out involving a reanalysis of the review's findings taking into account the uncertainty in the methods and the data. This helps to determine if the inferences of a systematic review change due to these uncertainties. In a primary study there may be withdrawals, so sensitivity analysis may involve repeating the analysis imputing the best or worst outcome for the missing observations or carrying forward the last outcome assessment. Also see Intention-to-treat analysis and Withdrawals.

Specificity (true negative rate) of a test The proportion of those subjects who really do not have disease who are correctly identified as such.

Standardised mean difference (SMD) Standardised difference in means is an effect measure for continuous data where studies have measured an outcome using different scales (e.g. pain may be measured in a variety of ways or assessment of depression on a variety of scales). In order to summarize such studies, it is necessary to standardize the results into a uniform scale. The mean difference is divided by an estimate of the within-group variance to produce a standardised value without any units. Also see Effect. (erroneously called standardised mean difference)

Strength of evidence The strength of evidence describes to the extent to which we can be confident that the estimate of an observed effect, i.e. the measure of association between interventions and outcomes obtained in the review, is correct for important questions.

Subgroup analysis Meta-analyses may be carried out in pre-specified subgroups of studies stratified according to differences in populations, interventions, outcomes and study designs. This allows reviewers to determine if the effects of an intervention vary between subgroups.

Summary receiver operating characteristics curve (SROC) A method of summarising the performance of a dichotomous test pooling 2x2 tables from multiple studies or multiple cut-off points. It takes into account the relation between sensitivity and specificity among the individual studies by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity).

Surrogate outcomes A substitute for direct measures of how patients feels, what their function is, or if they survive. They include physiologic variables (e.g. blood pressure for stroke or HbA1c for diabetic complications) or measures of subclinical disease (e.g. degree of atherosclerosis on coronary angiography for future heart attack). To be valid, the surrogate must be statistically correlated with the clinically relevant outcome but also capture the net effect of the intervention on outcomes. Many surrogates lack good evidence of validity.

Systematic error see Bias.

Systematic review (systematic overview) Research that summarises the evidence on a clearly formulated question using systematic and explicit methods to identify, select and appraise relevant primary studies, and to extract, collate and report their findings. By following this process it become a proper piece of research. It may or may not use statistical meta-analysis.

Theme An idea that is developed by the categorisation of qualitative research data under its heading. The large quantities of data produced by qualitative study are managed by the generation of themes and the coding of parts of the data to each theme. The perspectives of each research participant on each theme can then be compared and analysed.

Theory Abstract knowledge or reasoning as a way of explaining social relations. Theory may influence research (deduction), or research may lead to the development of theory (induction).

Trial see Clinical trial.

Triangulation Triangulation is the application and combination of several research methodologies in the study of the same phenomenon.

Validity (internal validity) The degree to which the results of a study are likely to approximate the 'truth' for the subjects recruited in a study, i.e. are the results free of bias? It refers to the integrity of the design and is a prerequisite for applicability (external validity) of a study's findings. Also see External validity.

Variance A statistical measure of variation measured in terms of the deviations of individual observations from the mean value. The inverse of variance of the observed individual effects is often used to weight studies in statistical analyses used in systematic reviews, e.g. meta-analysis, meta-regression and funnel plot analysis.

Withdrawals Participants or patients who do not fully comply with the intervention, cross over and receive an alternative intervention, choose to drop out, or are lost to follow up. If an adverse effect is the reason for withdrawal, this information can be used as an outcome measure. Also see Attrition bias, Intention-to-treat analysis and Sensitivity analysis.