**Abductive reasoning (abduction)** is logical thought process that moves from the specific to a conjectured explanation (‘best guess’) (Peirce, 1955). It is used to explain a specific observation/experience where there is no prior knowledge available; it can be intuitive and creative, often featured in theory development and used with other forms of reasoning. For example: a patient presents with a set of symptoms and the doctor makes an initial diagnosis (inferred through abduction); the doctor may then test that explanation by doing a diagnostic test. Also see **Inductive reasoning** and **Deductive reasoning**.

**Absolute benefit increase (ABI)** is the absolute difference in event rates between the experimental and control groups (i.e., ABI=EER-CER)

**Absolute risk** is the calculated or observed probability or risk that a population will be exposed to an event. For example, in the population of women taking a type of contraceptive pill the absolute risk of thrombosis is 1 in 7000. In terms of statistic literacy, absolute risk is easier to communicate than **Relative risk**.

**Absolute risk reduction (ARR)** is the absolute difference in event rates between the experimental and control groups (i.e., ARR=CER-EER)

**Adjusted analysis** is a statistical method, usually described in the data analysis section of an article that checks whether or not subgroup differences and predictions are the result of other important prognostic factor/s.

**Adoption**, in the context of evidence implementation, refers to clinicians’ commitment to and actual change in their practice (Rycroft-Malone & Bucknall, 2011).

**All or none study** a rare type of **case series study design** where all or none of the cases (participants) in a series who each have the same risk factor/s or prognostic factor/s experience the same outcome. It is rare because it is difficult to get a case series that is truly representative of the broader population. See **representative sample**.

**Allocation** is the manner that **randomisation** occurs and participants are assigned to a study group. Ideally the process of allocation will be concealed, for example by use of: central randomisation; sequentially numbered, opaque, sealed envelopes; sealed envelopes from a closed bag; numbered or coded bottles or containers; drugs prepared by the pharmacy; or other descriptions that contain elements convincing of concealment. Inadequate concealment may occur due to, for example: no concealment procedure; sealed envelopes that were not opaque; or other descriptions that contain elements not convincing of concealment (Health Information Research Unit, 2011)

**Analytic observational studies**, are **Epidemiological studies** where the researcher passively observes and measures the exposure or treatments of the groups. These studies all assess associations between exposures and outcomes; see **Analytic study designs**.
**Analytic study designs** attempt to quantify the relationship between two factors, i.e., the effect of (I) an intervention, issue or exposure on (O) an outcome. To quantify the effect, the rate of outcomes in (C) a comparison group and intervention or exposed group is also needed. Analytic study designs may be **observational analytic** or **experimental**.

**ANCOVA**: Abbreviation of ANalysis of COVAriance. This is a statistical procedure used to test mean differences among groups on a **dependent variable** and tries to ensure that other variables (covariates) are not influencing the apparent relationship.

**ANOVA**: Similar to ANCOVA and stands for ANalysis Of VAriance. This tests the mean difference between three or more groups and compares how much variability there is within the groups as well as between them.

**Audit trail**: In qualitative research the detail that indicates the way the researcher moved from individual quotes, or observations, to key categories used to make sense of the findings. This contributes to the credibility of the research.

**Auditability**: In qualitative research, the judgement that the researcher has provided sufficient detail to allow the reader to follow an audit trial and confirm the researcher’s conclusions.

**Background information** consists of any forms of published media, including textbooks and research-based evidence that is variously relevant to the topic but does not answer the information need; this type of evidence can provide definition to key search terms and/or prompt future research more specific to the information need (for example, animal research and laboratory studies have a place in the development of drugs). Background information can include anecdotal evidence and/or be heavily influenced by beliefs, opinions, or even politics.

**Bar graph**: a way of showing discrete (nominal or ordinal) numeric results that illustrates differences between variables or groups using separate blocks.

**Before-and-after designs** are where measures on an outcome are taken before and after the intervention is introduced to a series of people and are then compared (also known as Pre-test/post-test design); see **Case series design**

**Before-and-after study** (or **Pre-test/post-test Case series**) is where particular outcome measures are taken on a group of people (Case series) before they are exposed to the same intervention (factor under study), measures on the outcome are taken after the intervention is introduced and then compared.

**Best research evidence** is research evidence that is valid, reliable and statistically significant with a study design suited to providing clinically significant data of the type required for making decisions to the relevant to the information need (Hoffmann, Bennett, & Del Mar, 2013)

**Bias** consists of systematic errors in results due to sampling, measurement or analysis. Bias produces results that differ in a systematic manner from true values. Sample size has no effect on bias. Bias can vary in size and may operate in either direction; for example, in evidence about interventions it may cause over or under estimation of the **intervention effect**. Bias does not mean imprecision (see **Random error**). As bias is a systematic error, replication of a biased study will reach the same wrong answer on average. See **Selection bias, recall bias**

**Bimodal distribution**: A statistical description of the distribution of a variable where the data indicate there are two values that occur with equal frequency. Plotted on a graph, a bimodal distribution would be indicated by two peaks.
Binary outcomes / Binary data see Dichotomous outcomes / data

Blind, independent comparison in Diagnostic accuracy studies means that: a) the index test is independent of the reference standard test and; b) the assessor determining the results of the index test is blinded to the results of the reference standard test and vice versa.

Blinding (or Masking) is when clinicians, patients, data collectors and/or data analysts are unaware of who did or did not receive the study intervention. In reports bracketing is used to indicate those who are blinded e.g., “blinded [patients and outcome assessor initially]”. The use of ‘initially’ indicates that blinding was broken during the trial, for example, due to adverse events. If ‘unblinded’ is indicated, it means that no one in the trial was blinded so that everyone was aware of who received the intervention.

Boolean operators link identified search terms in a meaningful way: Using AND generally limits the search, results contain both terms (e.g., Insulin AND infusion will find material that contains both the words insulin and infusion, it will not find material that has only one or other term); Use of OR usually broadens the search and results contain one or other or all of the terms separated by OR (e.g., AIDS OR Acquired immune deficiency syndrome will find material that contains either or both of the words AIDS or Acquired immune deficiency syndrome); NOT limits the search and should be used with caution, results exclude the subsequent term (e.g., Insulin NOT infusion will find material that contains insulin but not the word infusion).

Bracketing as part of your search strategy: Brackets are used to group variants of one search term together; such bracketed groups can then be linked into a full search strategy (or search ‘string’) using Boolean operators.

Bracketing in qualitative research: researchers are encouraged to identify their own experiences or expectations that may influence their preconceived ideas about the study, and put aside or ‘bracket’ these so that they do not unduly influence interpretations of the findings. There is some controversy over the extent to which this is possible to achieve.

Case-control studies are observational analytic studies in which people who already have a specific condition (i.e., cases) are compared with an appropriately similar group of people who do not have the condition (i.e., controls). The researcher looks back (i.e. the study is ‘retrospective’); factors or exposures that might be associated with the condition (factor under study) are identified, often using medical records or patient recall for data collection.

Case reports are reports on the treatment of a single patient.

Case series design consists of a group (i.e., a single series) of people (i.e., cases) exposed to the same intervention (factor under study); Post-test is where only the outcomes after the intervention are recorded in the series of people, so no comparisons can be made; Pre-test/post-test is where measures on an outcome are taken before and after the intervention is introduced to a series of people and are then compared (also known as a Before-and-after study).

Causal relationship: The objective of experimental designs is to establish evidence of a causal relationship between an independent variable (the cause) and a dependent variable (the effect). It is confirmed by the statistical relationship between the two variables.

Causation statistics indicate a cause and effect relationship between variables, i.e. one event B is the result of another event A. The most effective way to establish causation is to undertake a controlled experimental study in which the sample participants are randomly allocated to either a control
group or a treatment group where a variable is manipulated (e.g. Randomised controlled trials). See inferential statistics

Central tendency is the estimated centre of a distribution; it is a statistical term for the result of a procedure that attempts to establish a typical single value from a set of numbers. Average is the common term, but there are three such measures: mean, mode and median. Each of these can produce a very different result because of the way in which they are calculated. See descriptive statistics

Chance is the random variation in results, see Random error

Change scores are calculated from the change in outcome over the study period

Chi-squared test (χ²): (Pronounced ‘ki-squared’): In statistics, a nonparametric test that seeks to establish if the difference between the observed results of a categorical variable (e.g. male, female) is statistically different from the expected value, and is unlikely to have happened by chance.

Clinical effect see Clinical significance

Clinical expertise is the ability to use professional skills and past experience to rapidly identify each patient’s unique health state and diagnosis, their individual risks and benefits of potential interventions, and their personal values and expectations (Straus, Glasziou, Richardson, & Haynes, 2011)

Clinical heterogeneity in a systematic review refers to differences between individual study findings associated with the participants, interventions or outcomes. Despite selection of similar studies there can still be substantial differences that mean study pooling is unfeasible.

Clinical practice guidelines are statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options (Institute of Medicine, 2011, p. 1)

Clinical significance (sometimes called clinical effect) is a judgement based on whether the results (e.g. intervention effect) are large enough to have an impact on patients i.e. create a clinically meaningful improvement. It is calculated using the no effect value, the smallest worthwhile effect and the confidence intervals (CI) of the intervention effect. If the entire CI is above the smallest worthwhile effect (and therefore the no effect value) then the intervention will be clinically significant. If the entire CI is below the smallest worthwhile effect (and possibly crossing the no effect value) then the intervention will not be clinically significant. When the CI spans the smallest worthwhile effect the intervention may be clinically significant but that conclusion cannot be made with certainty.

Cluster Trial design is a Randomised Controlled Trial where two or more clusters of people (such as townships or hospitals) are allocated to either a control group or to one or more intervention (issue/indicator/exposure) groups.

Cluster Trial design is an RCT: where two or more clusters of people (such as townships or hospitals) are all located to either a control group or to one or more intervention (or exposure) groups.

Coherency see Congruency

Cohort studies (may also be called epidemiological studies or longitudinal studies) identify a large group of patients who are already taking a treatment or have an exposure. This group is followed forward over time (i.e., Prospective cohort study) and then their outcomes are compared with a
similar group that has not been affected by the treatment or exposure being studied. Cohort studies are *observational analytic studies* and not as reliable as *randomised controlled trials*, since the two groups may differ in ways other than in the variable under study. See also *Retrospective cohort study*.

**Concepts** are abstract ideas or mental images of observed phenomena or a phenomenon (Chinn & Kramer, 2010). For example, the concept ‘dog’ is not specific to a particular dog, it classifies all types of four legged hairy animals who bark and can breed with each other as ‘dogs’. Concepts are different from *variables* because concepts provide a direct link back to reality or agreed-upon reality in terms of values (e.g. the concept ‘truth’) or emotions (e.g. the concept ‘happiness’) (Rycroft-Malone & Bucknall, 2011; Watt & van den Berg, 2002).

**Conditional probabilities** are also called *Relative frequencies* they are statistical measures of the number of events in a population comparing the number of events in another population; the measures express proportion that is often ‘normalised’ into a percentage (e.g. a proportion of 0.2 is 20%). Examples are *Sensitivity* are *Prevalence*. See *Natural frequencies* for how these proportions are expressed in plain language.

**Confidence interval (CI)** is a quantification of the range of uncertainty (Random error) in measurement; i.e. how confident the researchers are that the true value of the results lie within a certain value range. CI describes the accuracy or level of *Precision of the intervention effect* produced by the study sample relative to the whole population. CI is usually reported as 95% interval meaning that we can be 95% certain that the true average intervention effect for the whole population lies within that range.

**Confirmability** in qualitative research considers: to what extent analytic decision-making during analysis (interpretation) is reflexively acknowledged to arise from the researcher/s in addition to being shown to be grounded in the data and the conditions of the research (Liamputtong, 2013).

**Confounder/confounding variables** are a type of *extraneous variable* that systematically vary in similar ways to the *independent variable*, thereby obscuring the outcome (it is a systematic error that can be controlled by *randomisation*).

**Confounding** is an error in interpretation caused by *confounding variables* (or *confounder variables*), which are often unknown, being confused with the variable of interest. The best means of controlling confounding is *randomisation*. *Restriction* and *matching* can also control the effect of confounding.

**Congruency** means ‘agreement, harmony and/or compatibility’. In research congruency refers to the ‘match’ or ‘fit’ between different aspects of the research; for example, the match of research design to the methods, the match between research question and study design, or the fit between the design/methods and the discipline/scenario. Congruent research is *coherent*, in that there is a logic, clarity and consistency between all its parts. *Methodological congruency* is the fit between the individual assumptions of the methodology with the study design/methods.

**Consecutive cohort studies** are used in diagnostic accuracy studies; usually consist of a single cohort where all participants are recruited consecutively regardless of whether they have the condition being tested for. Participants will then have the *index test* followed by the *reference standard*, consecutively. Ideally test interpreters will be double blinded (i.e. all interpreters of either index or reference standard test are blinded to all other test results).
Consecutive recruitment is the ideally form of recruitment for Diagnostic test accuracy studies (Consecutive cohort studies). All participants should be recruited consecutively regardless of whether they have the condition being tested for.

Constant comparative method: In qualitative research, this is a method of analysing the data where comparisons are made with previously noted items or categories to ensure consistency in the method of analysis.

Constructs, like concepts are also mental images but they are built from the logical combination (using propositions) of concepts (Watt & van den Berg, 2002). Constructs are therefore more abstract than concepts. Sometimes the terms are used interchangeably (e.g., the concept/construct ‘expertise’) but the closer link to reality is maintained by concepts (e.g., the sub-concepts of ‘expertise’ are ‘formal education’ and ‘experience’).

Contextual evidence is a combination of patient values and patient circumstances (merged in some EBP contexts such as violence prevention) (Puddy & Wilkins, 2011)

Continuous outcomes / Continuous results / Continuous data are types of variable that lie along a continuum and have a potentially infinite value (e.g., weight, age, rating on a visual analogue scale)

Continuous variables produce quantitative data that has been measured using an interval scale, an index or a ration (i.e. interval data, ratio data). These numbers can be added or subtracted (and sometimes divided and multiplied). For example, weight in kilograms or temperature in Celsius or measurements from a standardised measurement tool.

Control event rate (CER) is the risk of the outcome of interest occurring in control group (eg: CER=a/(a+b) where a= Outcome present in control group and b= outcome not present in control group)

Correlational studies are observational analytic designs that measure the relationship – correlation – between two variables in the same participant group. These studies are cross-sectional designs that are analytic.

Correlation is a statistical technique that searches for a relationship between two variables in a study. A positive correlation means that as one variable increases, so does the other (e.g. height and weight); a negative correlation is where as one variable goes up, the other goes down (e.g. outdoor temperature and weight of clothing). The strength of a correlation is indicated by a correlation coefficient, which indicates how certain we can be that a clear pattern exists. (Rees, 2011)

Credibility in qualitative research considers: how well the researcher interpretation of data fits with what the participants have said; and, how believable the findings are (Liamputtong, 2013).

Critically appraised individual articles include a summary of the key methods and findings of the paper and provide a critique or commentary related to clinical practice.

Critically-appraised topics (CATS) are standardised one page summaries that present the best available evidence for a particular clinical question. Authors of critically-appraised topics evaluate and synthesise multiple research studies.

Critique/Critical analysis is a process that can be positive and negative where information is not merely relayed; information is interpreted and/or commented upon. In critical analysis citation/s from relevant literature (eg. on methods/methodology) are given to support statements made about the information that is critiqued.
**Crossover Trial** design is a **Randomised controlled trial** where each study participant has both therapies, e.g., they begin when they are randomised to treatment A and then at the crossover point treatment B is started. All subjects serve as their own controls and therefore error variance and sample size is reduced. Can only be used when the treatment outcome (e.g., symptoms) is reversible with time.

**Cross-sectional studies** (often **Surveys**) are studies where a group of people are assessed at a one point (or cross-section) in time and the data collected on exposures and outcomes relate to that point in time; these studies may be **Descriptive study design** or **Observational analytic study design**. Cross-sectional studies are useful for hypothesis-generation, to identify whether a **risk factor** is associated with a certain type of outcome, but more often than not (except when the exposure and outcome are stable e.g. genetic mutation and certain clinical symptoms) the causal link cannot be proven unless a time dimension is included (Merlin, Weston, & Tooher, 2009). One exception is when the exposure and outcome are stable e.g. genetic mutation and certain clinical symptoms. Cross-sectional studies are effective at quantifying **diagnostic test accuracy** and quantifying prevalence of conditions or risks.

**Cross-tabulation table** (also called contingency table) is a modification of the frequency table that shows a breakdown of the first variable (e.g. age) in the frequency table using another variable (e.g. smoking status).

**Cut-off** refers to the point, which may vary, at which a disease is ruled ‘in’ (present) or ‘out’ (absent). For example, hypertension can be defined using the cut-off point of BP above 120/80, in other circumstances hypertension might be defined as above 130/90. Where the cut-off point that rules a disease ‘in’ (present) or ‘out’ (absent) is variable, test characteristics will also vary such that there will be a trade-off between the test’s **sensitivity** and **specificity**.

**Deductive reasoning** is a logical thought process that moves from a more general theory or premise to the specific; if the premise (statement or proposition) is true then the reasoning is true. For example: all birds have feathers and owls are birds, therefore owls have feathers.

**Delimiters** act to filter the data and thereby limit the number of items returned in the search. Also, known as ‘filters’ or ‘limits’; delimiters are normally placed at the beginning of a search, they can then be refined as necessary. Choice of these filters should be based on the question and study type but they are also used differently in different databases. A filter of ‘publication’ or ‘material type’ (e.g. journals, conference papers, newspapers) is available on most databases. Some databases have filters for study design or evidence type; for example, when specifying forms of primary evidence such as ‘randomised controlled trial’ or ‘cohort study’ or for specifying ‘systematic review’ or other types of review. Databases usually allow you to sort findings by date or to restrict your search to certain date ranges. Many databases allow you to limit your search to English language.

**Dependability** in qualitative research considers: how congruent the methods are with each other, the researcher decisions and the overall research approach; and, how logical, traceable and clearly documented the chosen methods and researcher decisions are (Liamputtong, 2013).

**Dependent variable** is the variable that (supposedly) changes, relative to the intervention in analytic, experimental designs e.g., **Randomised control trials**; or the exposure in analytic observational study designs, e.g. **Cohort studies**. This variable is the outcome in any type of analytic design, i.e. the ‘O’ in a PICO question.

**Descriptive statistics** are numerical techniques (i.e. **univariate analysis**) used to describe and summarise the data e.g. summing the data, averaging, calculating percentages and frequency...
distributions. Tables, bar graphs, histograms and pie charts are variously used to assist this description.

**Descriptive study designs** do not quantify relationships but give a picture of what is happening in a population (see Epidemiology), e.g., the prevalence, incidence, or experience of a group. Descriptive studies might be called ‘non-analytic observational studies’; they measure the frequency of several factors, and hence the size of the problem.

**Diagnostic case control studies** compare **reference standard** results with **index test** results for two groups: one with disease/condition and one without, both diagnosed using the reference standard test.

**Diagnostic test accuracy** refers to the amount of agreement between the **index test** (i.e., the test under study) and the **reference standard** test (i.e. the ‘gold standard’ test) in terms of outcome measurement. Diagnostic accuracy can be expressed in many ways, including sensitivity and specificity, likelihood ratios, diagnostic odds ratio (OR), and the area under a receiver operator characteristic (ROC) curve.

**Diagnostic yield studies** identify the number of people who have the disease/condition by using the **index test** but no reference standard is used for comparison so such studies are low on the diagnosis hierarchy (Hoffmann et al., 2013).

**Dichotomous outcomes / Dichotomous data** (or **Binary outcomes/data**) are types of variable that can take only one of two values (e.g., migraine or no migraine, return to work or no return to work, dead or alive)

**Diffusion**, in the context of evidence implementation, refers to the passive, unaided distribution of information; i.e., diffusion occurs naturally through clinicians’ adoption of policies, procedures, and practices (Rycroft-Malone & Bucknall, 2011).

**Discrete variables** (also called **Categorical variables**) produce qualitative information (using quantitative data collection methods); this ordinal or nominal data is numerically coded in some way.

**Dispersion** is the way values spread from the **central tendency**; it is described as either the range from high to low or as the **standard deviation**.

**Dissemination**, in the context of evidence implementation, refers to the active communication of information to a selected target audience of clinicians to improve their knowledge or skills (Rycroft-Malone & Bucknall, 2011).

**Distribution** is a summary of frequency, perhaps shown as a frequency distribution table or as a bar chart or histogram.

**Effect size** is a quantified measure of the size of the difference between two groups (or two variables) e.g. intervention/exposure group and the control/matched group.

**Emic** refers to an approach to data analysis that uses the participants’ perspective (the researcher is trying NOT to put their own interpretations on the data). See also **Etic**, note that sometime both perspectives are used.

**Empiricism** is the theory that all knowledge is based on experiences that arise from the senses (Blackburn, 2008)

**End scores** are those scores recorded when intervention is over i.e., at the end of the study period.
**Epidemiology** concerns studies at the population level; the unit of study is aggregate groups of people. The epidemiologist is interested in causation and population level interactions. Epidemiologic studies investigate: occurrences of events affecting health status; identification and characterisation of risk factors for adverse health events; identification and modes of action of interventions for enhancing health status. **Epidemiological studies** may be a **Descriptive study design** or **Observational analytic study design**.

**Epidemiological studies** see **Epidemiology**

**Epistemology** is the study of how knowledge is acquired and justified; it describes how the researcher gets knew knowledge and how new things are discovered. Concerned with distinguishing between opinion, belief and prejudice by establishing secure foundations for knowledge in particular regarding the limits, truth and methods of how that knowledge came about (Scott & Marshall, 2009).

**Ethics** is concerned with moral principles, values and what ought to be done; essentially with relationships between people and how they live (Johnstone, 2009).

**Etic** refers to an approach to data analysis that uses an outsider’s perspective (the researcher is using their knowledge etc to interpret the data). See also **Emic**, note that sometime both perspectives are used.

**Evidence implementation** is the active and systematic integration of information into practice; it involves identifying barriers to change, targeting effective communication strategies to address barriers and using administrative and educational techniques to increase effectiveness (Rycroft-Malone & Bucknall, 2011)

**Evidence-based practice** is an integration of four key factors: the best research evidence, clinical expertise, patient values and circumstances, and practice context (Hoffmann et al., 2013; Straus et al., 2011)

**Experiential evidence** is a combination of clinical expertise and practice context (merged in some EBP contexts such as violence prevention) (Puddy & Wilkins, 2011)

**Experiment event rate (EER)** is the risk of the outcome of interest in occurring experiment group (eg: EER=c/(c+d) where c= Outcome present in experimental group and d=outcome not present in experimental group)

**Experimental study designs** are **Analytic studies** where the researcher actively manipulates the exposure or uses an intervention. Subjects are randomly allocated to one or more intervention (or exposure) groups or a control group, they are then followed up under carefully controlled conditions. See **Randomised controlled trials**

**Expert opinions** refers to the views of professionals who have expertise in a particular form of practice or field of inquiry, such as clinical practice or research methodology (International Centre for Allied Health Evidence, 2014).

**Explanatory clinical trials** (Randomised controlled trials) are planned experiments undertaken in controlled conditions that can provide evidence of cause and effect. See **Randomised controlled trials**

**External validity** is the extent to which results are generalizable or applicable to a pre-defined population.
**Extraneous variables** are any variables that are not being studied but that could influence the research outcome—these variables increase error and will ideally be controlled via the study design.

**Factorial design** is a type of [Randomised controlled trial](#) where two or more ‘factors’ are evaluated simultaneously: both separately and in combination as well as compared to the control. ‘Factor’ refers to an independent variable or treatment, manipulated by the researcher. In factorial designs, different combinations of one or more factors (treatments) are used to explore outcomes. For example, a trial of the combined or separate use of an opiate and nonsteroidal anti-inflammatory (NSAID) drugs for treating cancer pain might have groups with/without: opiate alone, NSAID alone, opiate and NSAID in combination, and/or with placebo.

**Focused critical literature reviews** in the context of research, are a literature reviews that are focused on critiquing the research directly relevant to the research question.

**Forest plots** are a pictorial representation of the amount of variation in the studies included in meta-analysis. The precision (confidence interval) of each study is depicted by the horizontal line with a symbol on that line representing the intervention effect. The vertical line is the line of no effect value (where treatment and control are equal). The value of the no effect line is one (1) for dichotomous data and zero (0) for continuous data. When the horizontal line crosses the vertical line, there is no significant difference between treatment and control groups. Pooled data is usually represented with a diamond.

**Frameworks** are organisational structures. A research framework refers to the organisational structure of the whole or part of a research project. The research process is really a framework depicting how a research project generally proceeds. Theoretical frameworks are organisational structures based on one or more theories.

**Frequency distribution**: Lists the numbers and percentages of each sub-group of the main group; usually shown in a frequency table which can be used to generate e.g. pie charts and histograms.

**Funnel plots** are a graphical tool that estimates the level of publication bias in the literature included in a systematic review. It consists of a scatter plot where each dot represents one of the selected studies plotted as per the study’s intervention effect size on the x-axis against a measure of the study’s sample size on the y-axis. The ideal funnel plot will result in a symmetrical inverted funnel; asymmetrical funnel plots suggest that some studies have been missed (usually smaller ones showing no effect).

**Hawthorne effect** is a recognised phenomenon where people enrolled in medical trials have improved general health status over those people who are not part of a study. The difference is presumed to be from the psychological boost imparted through participation in a study, along with the greater focus on overall health that participants in a study receive.

**Health research** is a planned, systematic process that generates new knowledge and refines and/or validates existing knowledge in such a way that health problems, practices and/or outcomes are influenced either directly or indirectly (Minichiello, Sullivan, Greenwood, & Axford, 2004).

**Heterogeneous** refers generally to a diverse set of elements. In systematic reviews heterogeneity refers to a set of studies with differences between participants, interventions, and measurement of outcomes. See Clinical heterogeneity, Methodological heterogeneity and Statistical heterogeneity.

**Histogram**: a way of showing continuous numeric results using a line drawing in the shape of a series of touching blocks.
Historical control study compares outcomes for a prospectively collected group of people exposed to the intervention (factor under study) with either (1) the outcomes of people treated at the same institution prior to the introduction of the intervention (i.e. control group/usual care), or (2) the outcomes of a previously published series of people undergoing the alternate or control intervention.

Homogenous refers generally to a similar set of elements. In systematic reviews homogeneity refers to similarities between participants, interventions, and measurement of outcomes across a set of studies.

Hypothesis refers to a prediction about the nature and direction of the relationship between the study variables or study groups (Thomas & Hodges, 2010). The null hypothesis assumes that the prediction is not correct; for example, where the hypothesis predicts a difference between two groups, the null hypothesis says there will be no difference or zero effect. The hypothesis is sometimes called the alternative hypothesis as it is alternative to the null hypothesis.

Implementation, see Evidence implementation

Inception cohort is a group of people consistently recruited at a common time early in the development of a specific disease/condition; this is a strategy for dealing with selection bias in cohort studies (e.g. prognostic studies).

Incorporation bias occurs in diagnostic accuracy studies when the index test (the test being examined) forms a part of the comparison reference standard; it can lead to overestimation of diagnostic accuracy.

Independent variable is the variable that refers to the intervention in analytic, experimental designs e.g., Randomised control trials; or the exposure in analytic observational study designs, e.g. Cohort studies. This variable is the ‘I’ in a PICO question.

Index test refers to the test being examined in diagnostic accuracy studies; it is ideally compared with a reference standard i.e. the most accurate ‘gold standard’ test available.

Inductive reasoning is a logical thought process that moves from the specific to a more general proposition or theory (which may or may not be true). For example: this bird has feathers, therefore every bird must have feathers.

Inferential statistics are statistical tests used to compare qualities of groups with each other and to make inferences about the overall population based on the study sample. They are used to calculate the probability that the outcomes observed in a particular study sample could extend to the overall population. Inferential statistics is one of the two main categories of statistical analysis that use the numerical results of a study as the basis of inferences to a wider group. The second main form of statistics is descriptive statistics, which provides a numeric summary of the situation found in a study.

Intention-to-treat analysis refers to when participant data is analysed in the group that corresponds with how they were intended to be treated, not how they were actually treated (i.e., if a participant was allocated to the intervention group but for some reason did not have they intervention, they should still be counted as part of the intervention group). Intention-to-treat is important because it preserves the value of randomisation.
Internal validity is concerned with random variations in results (chance), systematic errors in results due to sampling, measurement or analysis (bias), and error in interpretation due to confounding variables being confused with the variables of interest (confounding).

Interrupted time series with a control group are studies where trends in an outcome or disease are measured over multiple time points before and after the intervention (factor under study) is introduced to a group of people, and then compared to the outcomes at the same time points for a group of people that do not receive the intervention (factor under study).

Interrupted time series without a parallel control group are studies where trends in an outcome or disease are measured over multiple time points before and after the intervention (factor under study) is introduced to a group of people, and then outcomes are compared within the study population rather than a separate (parallel) control group.

Interval data is ranked ordinal data that is arranged on a linear scale at equal intervals (may have positive and negative values), e.g. the Celsius scale of measuring temperature by degrees – has a zero and negative values.

Intervention effect is the amount of difference that an intervention makes to an outcome. It is reported differently depending on whether outcome variables are continuous or dichotomous, i.e. the initial research variables and therefore the outcomes lie along a continuum (e.g. weight, age, rating on a visual analogue scale) or are dichotomous (e.g. results are compare the two groups and can be used to determine). Continuous outcomes are reported in terms of Mean, Median and Standard deviation. Dichotomous outcomes report, e.g. Relative risk (RR), Relative benefit increase (RBI), Relative risk reduction (RRR), Odds ratio (OR) and Number needed to treat (NNT).

Level of significance: The p value or probability level in a study that indicates the extent to which the results could have happened by chance (Random error); it is the probability of making a Type I error (finding a relationship between groups/variables when that relationship does not exist). The minimum level is usually set at ‘p < 0.05’ which means that, statistically, there is less than five in a hundred chance of being wrong if the researcher maintained there was a relationship between the independent and dependent variables in the study.

Logic refers to reasoning such as deductive, inductive and abductive reasoning; it is one of the modern branches of philosophy.

Longitudinal study designs are where data is gathered on the same group of people multiple times; depending on the individual research design, intervals between data collections may be short or long, regular or irregular. These designs help to determine the pattern of change over time.

Masking see Blinding.

Matching may control the effect of confounding in a case-control study. Here subjects are chosen in such a way that potential confounders are distributed in an identical manner in each study group, hence the term ‘matched’ groups.

Mean is the average of the numbers calculated by adding all the numbers and dividing by the total number of numbers (e.g., there are five numbers: 18, 26, 14, 6, 12 the total comes to 76/5 = 15.2 which is the mean).

Median is the middle number of a list of numbers after they have been sorted numerically (e.g., there are five numbers: 18, 26, 14, 6, 12; when sorted they are 26, 18, 14, 12, 6, the middle number is 14 and that is the median.)
MeSH stands for ‘Medical Subject Headings’ – these are the database specific headings used to search on Medline and some other databases.

**Meta-analysis** usually occurs within the context of a **systematic review**. Meta-analysis involves the thorough examination of all the valid primary quantitative studies on a topic and mathematically combines the results using a pre-defined, accepted methodology; results are reported as if it were one large study. See **Qualitative evidence synthesis** for the similar qualitative process.

**Meta-regression** is a technique that explores which types of patient-specific factors or study design factors contribute to heterogeneity, see **heterogenous**.

**Meta-synthesis**, also called **Qualitative evidence synthesis**, usually occurs within the context of a **systematic review**; involves the thorough examination of valid primary qualitative studies, findings are pooled and reinterpreted using an accepted methodology to create new understandings, theories, practice or policy recommendations which are reported as it one large study. See **Meta-analysis** for the similar quantitative process.

**Methodological congruency** is the fit between the individual assumptions of the methodology with the study design/methods. See **congruency**.

**Methodological heterogeneity** in a systematic review refers to differences in the way that individual studies were conducted – such as between study design or risk of bias.

**Methodology** is the study of ontology and epistemology as applied to research; it is principally concerned with the wider philosophy of how research is undertaken including issues of validity (Scott & Marshall, 2009). Methodology in an individual research project describes the systematic ways that the new knowledge will be discovered and analysed: namely, the specific philosophical and ethical framework (e.g., assumptions) that underlies the choice of research design and methods that are used to produce and validate the ‘new’ knowledge (i.e., results/interpretation).

**Methods** are the individual techniques and strategies used systematically by the researcher to collect data, and to analyse and interpret that data, and thereby acquire/create new knowledge.

**Mixed method** studies are individual studies that use both quantitative and qualitative methods; mixed method **systematic reviews** use data from both types of study, quantitative and qualitative design. Note, sometimes the term ‘mixed method’ is used to refer to studies that use more than one different type of method despite using either all quantitative methods or qualitative methods.

**Mode** is the value (or characteristic if a **discrete variable**) that occurs most frequently in a data set; it may be considered the most common or popular.

**Models** generally take a narrow scope in representing an individual concept or construct. Concepts within models are ideally well defined and inter-relationships between the concepts are clearly specified. Models attempt to objectify the concept they represent; they present a representation of reality (Rycroft-Malone & Bucknall, 2011).

**Natural frequencies** use plain language to express proportions; i.e. stating the actual numbers of people experiencing an event within one group which can then be compared with actual numbers of people experiencing an event in another group (often a subgroup). For example, a proportion of 0.2 (20%) explained as 2 in 10 people (or 20 in 100, or 200 in 1000). See **Conditional probabilities**.

**Negative likelihood ratio (LR-)** a ratio of negative probabilities; the calculation is (False Negative/all people in the sample who have the disease)/(True Negative/all people in the sample who do not
have the disease), i.e., \((FN/(TP+FN))/(TN/(TN+FP))\). It may also be calculated as (100-

sensitivity)/specificity, when only these results are in the article. When the \(LR>0.5\) the test helps to

rule that the disease is truly not present and \(LR>0.1\) indicates that the test is very good at ruling the
disease as truly not present.

**Negative predictive value** is the probability that a person does not have the disease if their test
result is negative; it is calculated by dividing all the true negative (TN) by the true negative (TN) plus
the false negative (FN), i.e. \(TN/(TN+FN)\). The closer the result is to 100% the most probable that the
test is accurate and the person does not have the disease. The complement of this value (i.e.,
\(FN/(FN+TN)\)) is the **Post-test probability of a negative test**.

**No effect value** refers to there being no effect between groups. When calculating the effect size for
continuous outcomes (i.e. using means or medians) it is zero. For **dichotomous outcomes**, the value
is zero for RRR and ARR but one for RR and OR. CI intervals that do not include the ‘no effect value’
are statistically significant.

**Nominal data** is data that is measured only in terms of whether it belongs to a certain category or
subcategory. There are three types of nominal data: constant (present/not present), dichotomous
(yes/no), non-dichotomous (e.g. drizzle, showers, rain, snow, sleet, hail). Nominal data is
quantitative data that from **discrete variables**.

**Non-analytic observational studies** see **Descriptive studies**

**Nonparametric tests** are statistical tests used to analyse data that do not conform to the normal
distribution; they can therefore establish the likelihood of relationships amongst data that do not
require the strict criteria of parametric statistics. This makes them easier to apply but the results are
not as widely accepted or as accurate as those in the parametric category. Nonparametric tests
make no assumptions because there is no fixed knowledge of factors pertaining to the study
population (e.g., mann-Whitney \(U\) test, rank sum and Kruskal-Wallis one-way ANOVA test, Wilcoxon
test, Friedman two-way ANOVA).

**Nonparametric tests** are tests that make no assumptions because there is no particular, fixed
knowledge of factors pertaining to the study population (e.g., mann-Whitney, rank sum and Kruskal-
Wallis tests).

**Non-positivist paradigms**, a collection of paradigms with alternative assumptions to positivism:
‘empirical’ is more than just data that can/cannot be verified by the senses; the quality of an
experience is at least as important, than the quantity of people who have an experience. Non-
positivism aims primarily to enhance human understanding (interpretivist/constructivist paradigms)
or to prompt human emancipation (Habermas, 1971). Non-positivist researchers have more
subjective values of truth including an assumption that truth changes with context and creates
multiple true realities. The non-positivist paradigms may include intuitive, tacit ways of knowing,
emotional responses, cognitive responses and bodily responses to experience. Non-positivism
maintains that both methodological and researcher values should clearly be identified and they
should be evident in research reports. See **Positivist paradigm**

**Non-probability sampling** refers to sampling where the probability that each member will be
included in the sample cannot be specified. The main advantages of non-probability sampling are
convenience and cost. However, with non-probability samples, probability statements about sample
statistics cannot be made. For example, it is not possible to compute a confidence interval for an
estimation problem.
Non-randomised, quasi-experimental design are carefully planned clinical experiments where the sample population (e.g., people, a cluster of people) are allocated to the intervention (or exposure) group/s or control group, using a non-random method (such as patient or clinician preference/availability) and the outcomes from each group are compared. Includes Controlled before-and after studies where outcome measurements are taken before and after the intervention is introduced, and compared at the same time point to outcome measures in the control group.

Non-significant result: This does not mean that the results of the research are not important; it means that in the case of a randomised control trial the ‘p value’ is too large to rule out the element of chance and therefore the null-hypothesis must be accepted. In other words, there was no real difference between the results of the groups involved.

Normal distribution refers to a naturally occurring symmetrical pattern of variation that, when interval data are graphed, looks like an inverted bell-shaped curve where most data are close to the mean, and others are in equal proportion on either side of the mean. In a normal distribution, the mean, mode and median all have the same value and graphically would be a single line dropping down from the apex of the frequency curve. A normal distribution allows the use of parametric tests, making this an important statistical concept.

Null hypothesis assumes that the hypothesis is not correct; for example, where the hypothesis predicts a difference between two groups, the null hypothesis says there will be no difference or zero effect.

Number needed to treat (NNT) is the number of people needed to be treated to achieve the event of interest once (ie. NNT = 1/ARR or 1/ABI when the ARR or ABI is a decimal). Or, NNT=100/ARR or 100/ABI if the ARR or ABI is a percentage. The NNT is rounded to next whole number and should be accompanied by its 95% confidence interval (CI).

Observational analytic studies, are Epidemiological studies where the researcher passively observes and measures the exposure or treatments of the groups. These studies all assess associations between exposures and outcomes; see Analytic study designs.

Odds ratio (OR) measures the relationship between exposure and outcome; it shows the likelihood of a disease given an event (e.g. exposure) compared to the likelihood of the disease when the event (e.g. exposure) has not occurred. It is calculated by dividing the event rate by the non-event rate for each group (i.e., OR = [EER/(1-EER)] / [CER(1-CER)]). An OR=1 means there is no effect between the groups.

Ontology is the study of being or what exists. Said differently, the study of reality including beliefs about reality (Blackburn, 2008).

Ordinal data is nominal data that can be ranked in some way from first to last, e.g. the grading system of fail, pass, credit, distinction and high distinction. Ordinal data is quantitative data that from discrete variables.

Original research/original studies – see primary evidence

Outliers: In a set of numeric measures, these are the values at the extreme ends of a distribution that are not necessarily typical of the others in the set, e.g. those untypically old or young in a group.

p value is a value ranging from 0 to 1 related to the probability that the results are due only to chance (Random error), i.e. the probability of making a Type I error (finding a relationship between groups/variables when that relationship does not exist). If p < 0.05 the probability of the result...
occurring by chance is less than a one chance in twenty or 5%. The *p value* provides an indicator that the researcher has statistically tested the results to establish the ‘probability’ that the differences could be due to chance, and not the intervention. The most common values used to indicate that the results are unlikely to have happened by chance are, in increasing size of certainty that the results are real differences, $p < 0.05$, 0.01 and 0.001.

**Paradigms** are sets of values, beliefs, assumptions and practices that are shared by a group. In science a paradigm is seen as a framework of research outcomes and theories that inform further research and theorising (Blackburn, 2008; Kuhn, 1970). See **Positivist paradigm** and **Non-positivist paradigms**

**Parallel Trial design** is an RCT: where each participant is allocated to either a control group or one or more intervention (or exposure) groups; the groups then run approximately simultaneously i.e., in parallel.

**Parametric tests** are tests that make assumptions based on knowledge of particular, fixed factors about the relevant study population (e.g. t-test, f-test, z-test, ANOVA).

**Patient circumstances** are the individual situation or clinical state of the person (Straus et al., 2011)

**Patient values** are the unique preferences, concerns and expectations that each person brings to an encounter and which requires integration into decision-making in order for their decisions to be beneficial (Straus et al., 2011)

**Patient-reported outcomes (PROs)** are an outcome reported directly by patients themselves and not interpreted by an observer; PROs may include patient assessments of health status, quality of life, satisfaction with care or symptoms, or patient-reported adherence to medication (Calvert, Blazeby, Altman, & et al., 2013). **PROMs** are the tools, usually self-contained questionnaires, used to measure PROs (Meadows, 2011).

**Philosophy** is broadly defined as ‘the love of knowledge, the pursuit of wisdom’. Philosophy is the advanced study of fundamental questions about existence, knowledge, reasoning, morals, politics and/or aesthetics. The main branches of philosophy today are: **ontology, epistemology, logic and ethics** (Blackburn, 2008; Teichman & Evans, 1999).

**Pie charts**: show proportions of subgroups in a group by dividing portions of the pie according to the percent of data in each subgroup.

**Placebo effect** is the phenomenon where some people experience some type of benefit after the administration of a neutral substance. In clinical trials the placebo effect can mean that people receiving a neutral substance do better in the course of disease than do people in the general population, even if they do not do as well as people receiving the therapeutic drug under study.

**Placebo** is an intervention that appears the same as the intervention (or exposure) being studied but contains no active ingredients

**Population** refers to the complete set of observations a researcher is interested in, whereas the **sample** that is studied is a subset of that population. A population can be defined in a manner convenient for a researcher. For example, one could define a population as all women hospitalised with Asthma in the past 3 months at XXX hospital, a different population is the set of all women hospitalised with Asthma in the past 3 months in Australia; the first population is actually a subset or sample of the second population.
**Positive likelihood ratio (LR+)** is a ratio of positive probabilities; the calculation is \((\text{True Positives/} \text{all people in the sample who have the disease})/(\text{False Positives/} \text{all people in the sample who do not have the disease})\), i.e., \((\text{TP}/(\text{TP+FN})]/(\text{FP}/(\text{TN+FP}))\). It may also be calculated as sensitivity/(100-specificity) when only these results are in the article. When the LR+ > 2 the test helps to rule that the disease is truly present and LR+ >10 indicates that the test is very good at ruling the disease as truly present.

**Positivist paradigm**, focuses on predicting and controlling the natural and social environment (Habermas, 1971). Positivism, also called logico-empiricism, is a paradigm that assumes single definitive truths can be found with the objective application of logical theory to empirical fact (Fahy & Harrison, 2005; Parratt, 2010). Empirical, for positivists, means data that is able to be verified by sense experience (Blackburn, 2008). Positivism has dominated scientific research, resulting in profound technological advances. Positivist research on human experience involves reducing experience to observable behaviours studied as discreet, decontextualised units; that means positivism’s successes in terms of understanding and/or controlling human activity are not as profound. Positivism’s objectivity means that values within positivism are generally denied and hidden. Research papers from the positivist paradigm often do not elucidate methodology beyond the methods. See **non-positivist paradigms, scientific method**

**Post-test design** is where only the outcomes after the intervention are recorded in a series of people, so no comparisons can be made; see **Case series design**

**Post-test probability of a negative test** or the Complement of the negative predictive value is the probability that a person has the disease even if their test result is negative; it is calculated by dividing all the false negative (FN) by the false negative (FN) plus the true negative (TN), i.e. \(\text{FN}/(\text{FN+TN})\). The closer the result is to 0% the most probable that the test is accurate and the person has the disease.

**Post-test probability of a positive test** or the **Positive predictive value** is the probability that a person has the disease if their test result is positive; it is calculated by dividing all the true positives (TP) by the true positives (TP) plus the false positives (FP), i.e. \(\text{TP}/(\text{TP+FP})\). The closer the result is to 100% the most probable that the test is accurate and the person has the disease.

**Power** is the probability of **NOT** committing a **Type II error**. At a given significance level, the power of the test is increased by having a larger **sample size**. The minimum accepted level is 80%, meaning that there a chance of eight in ten that a difference of the specified effect size will be detected. See **Statistical Power**

**Practice context** is the clinical or work setting (Hoffmann et al., 2013)

**Practice-based evidence** takes varying forms including: professional experience and theoretical knowledge (e.g. past education, continuing professional development, clinical experience); practitioner derived reflective practice (e.g., following Gibbs’ (1988) reflective cycle or similar); clinical audit or original practice-based research; cultural practices that have been considered effective through community consensus (Lieberman et al., 2010; Martinez, 2008; Payne, n.d.). Practice-based evidence often becomes the prompt to undertake original research where data is drawn from practice.

**Pragmatic clinical trials** are a form of **Randomised controlled trial** designed to study the effectiveness of an intervention when used in normal practice; these trials help clinicians choose between options for care (Zwarenstein et al., 2008). See **Randomised controlled trials**
**Precision of intervention effect** indicates the accuracy of the **Intervention effect** for the study sample relative to the entire population. Usually shown by the **confidence interval** (CI) which is the range within which the true effect lies for the whole population.

**Pre-test/post-test design** where measures on an outcome are taken before and after the intervention is introduced to a series of people and are then compared (also known as a **Before-and-after study**); see **Case series design**.

**Prevalence** or **pre-test probability of having a disease** is the likelihood of a person having the disease prior to having the **index test**. It is calculated by dividing the number of people who are already known to have the disease by all the people who could have the disease whether they have the disease or don’t have it. For example, a sample of 50 people have the symptoms of the disease, but only 15 of those people truly have the disease: so 15/50=0.3 means the prevalence in the sample is 30%.

**Primary evidence** are reports of original research studies written by the researcher/s who undertook the research. Generally, such reports are published as articles in peer-reviewed journals. These reports should include detail about why the research was done (Background and/or Literature Review), how it was undertaken (Study Design, Methods), what was found (Data and Findings or Results) and what those findings mean (Discussion and Conclusions). Although these reports may be long, there should be an Abstract that clearly identifies the main parts of the study.

**Probability** is the likelihood that a given event will occur. In research, probability fundamentally involves asking ‘Given a finding from research, what is the probability that such a finding could occur by mere chance (random error)?’ Statistically probability is usually measured as the significance level (**p value**) and/or described using the **Confidence Interval**. **Frequency probability** is the probability that a randomly chosen circumstance will correlate with our expectation of it being chosen. **Model-based probability** is the probability of a proposed model of events occurring, asking for example, ‘what is the probability of a person having Alzheimer’s disease given that one of the person’s parents had it’. **Subjective probability** is based on a personal or professional judgment of whether an outcome is likely to occur, asking for example ‘what is the probability that a person has tuberculosis given that they have a positive skin test (PPD test)?’

**Probability sampling** is where every member of the population has a known probability of being included in the sample. Probability sampling enables computation of probability statements about sample statistics, e.g. confidence interval for an estimation problem.

**Prognostic factors** are any characteristic/s that could influence the outcome of interest such that its association is strong enough to accurately predict how the outcome develops. They may be co-morbidities such as obesity, disease specific such as severity of a disease and/or demographic such as gender. Prognostic factor/s are often included in the **P** of a **PICO/PIO** question. Prognostic factors are not the same as **risk factors** which are associated with the risk of development of a disease (not necessarily its outcome).

**Propositions** are statements used in theories to explain how particular **concepts** are related to particular other concepts and in what way (Rycroft-Malone & Bucknall, 2011).

**Prospective** designs look forwards in time, i.e. when participants are recruited to the study, the problem/situation being investigated has not yet occurred (or the treatment/intervention has not yet been introduced). Data is collected in the future relative to when participants joined the study.
**Protocol**: similar to a research proposal, a protocol is a clear plan, or set of steps to be followed in a study; should be used in all clinical trials, including Randomised controlled trials, and Systematic reviews.

**Pseudo-randomised controlled trials** are like RCTs but the sample population (e.g. people, a cluster of people) are allocated to the intervention (or exposure) group/s or control group, using a pseudo-random method (such as alternate allocation, allocation by days of the week or odd-even study numbers) and the outcomes from each group are compared.

**Publication bias** or **Reporting bias** is a type of bias created by the tendency to publish research reports that are statistically significant or otherwise support the preferred result. This form of bias also occurs when published studies present biased results, such as studies that only include outcomes where statistically significant differences were found. Failure to include unpublished studies in systematic reviews can bias results e.g., the true effect of an intervention may be overestimated. Publication bias may be represented graphically using a funnel plot or statistical methods.

**Purposeful sampling** involves selection of the sample based on participant characteristics that are likely to yield high quality data directly relevant to the research question. Some of the ways that such a sample may be chosen include: convenience (e.g. attendees at a clinic accessible to the researcher); snowballing, where new participants are referred by existing participants; and, maximum variation where participants are chosen for how different they are from each other (i.e. the extremes of the population are targeted during recruitment).

**Qualitative evidence synthesis**, may be called meta-synthesis, usually occurs within the context of a systematic review; involves the thorough examination of valid primary qualitative studies, findings are pooled and reinterpreted using an accepted methodology to create new understandings, theories, practice or policy recommendations which are reported as it one large study. See Meta-analysis for the similar quantitative process.

**Qualitative research** is defined by the Cochrane Collaboration as a study that uses a qualitative method of data collection and analysis (Higgins & Green, 2011). Such studies focus on people’s experiences, attitudes and beliefs and their perspectives of a situation or their understanding of an issue. The data are generally words (not numbers) and the words are analysed in non-statistical ways. Qualitative studies tend to use inductive reasoning to generate theories. CAUTION: studies about the quality of experiences/attitudes and/or perspectives often use surveys that quantify responses– in these studies the data are numbers, not words, so such studies are quantitative.

**Quantitative research** are studies that focus on the measurement of observations, usually in controlled situations. They aim to quantify variations in a phenomenon – measured as ‘variables’. The data are numbers that are analysed statistically. Quantitative studies tend to use deductive reasoning to test theories. NOTE: in quantitative research, feelings, judgements and other perceptions of quality can be ‘operationalised’ as variables and measured numerically focuses on measurement of observations in controlled situations. The data are generally numbers analysed statistically.

**Random error** is the random variation in results that is not due to systematic bias i.e., random error does not result from systematic processes that lead to systematic inaccuracies. Random error occurs because research investigates only a sample of the whole population; this means that multiple replications of the same study will produce different results as the sample is always slightly different. Chance and random error are synonymous and the direction of any error is unpredictable (see Type I
error and Type II error). In research, the possibility of random error is dealt with using confidence levels and testing for statistical power and statistical significance (p value).

Random sampling refers to the process of selecting a subset of a population for the purposes of statistical inference (using inferential statistics). Random sampling means that every member of the population is equally likely to be chosen.

Randomisation is the random assignment of people to the intervention group and the control group. Different designs for randomisation exist including simple, cluster, stratified and block randomisation. Randomisation is the best way to control for confounding. Through randomisation, potential confounders, known and unknown, as well as bias, are distributed equally between the exposed and control group. The goal is to end up with equal presence of the confounder in the study and comparison groups. The confounder(s) (and any source of bias as well) are still there, but their effects are cancelled out by being equally present in both groups, so that if a difference is observed between the two groups, the investigator can be very confident that the difference is the result of the exposure itself, not the confounder, i.e. ensuring that all the study participants are randomly exposed to various confounding factors.

Randomised controlled trials (RCTs) are carefully planned clinical experiments that introduce a treatment or exposure on real patients. Subjects are randomly allocated to one or more intervention (or exposure) groups or a control group, they are then followed up under carefully controlled conditions. RCTs may take a more explanatory attitude or a more pragmatic one. Explanatory clinical trials are planned experiments undertaken in controlled conditions that can provide evidence of cause and effect. Pragmatic clinical trials are designed to study the effectiveness of an intervention when used in normal practice; these trials help clinicians choose between options for care (Zwarenstein et al., 2008).

Ratio data is interval data that begins at zero as a fixed point; usually not linear, the intervals can be explained by mathematical operation; e.g. the growth of a bacterial population

Rationalism the theory that reason, not experience, underpins and provides certainty in knowledge.

Recall bias occurs when a study relies on participants to look back in time and identify events that may or may not be related to the outcome being studied. An example of this could be a study investigating a possible link between pesticide exposure and birth defects. If the study was in a community which already believed this was the case, it is quite likely that parents of children with birth defects may look back more carefully over past exposures than parents of healthy children. They may remember a minor exposure which other parents may well have forgotten. This too could result in an association being shown when none existed.

Receiver operating characteristic (ROC) curve is a graph that shows how sensitivity and specificity change according to changes in the cut-off point, see example in textbook (Hoffmann et al., 2013, p. 150)

Reference standard is used in diagnostic accuracy studies when referring to the most accurate ‘gold standard’ test available to compare with the index test (i.e., the test being examined). The reference standard or ‘gold standard’ test is ideally an established diagnostic test with defined performance characteristics. Choosing the ‘gold standard’ is not always straightforward as the best reference standard may be invasive (e.g. pulmonary angiography), too costly or time consuming (e.g. long-term follow up after the occurrence of syncope or seizures) or only be possible posthumously (e.g. neuropathic examination for Alzheimer’s disease).
**Regression** in terms of statistics refers to "prediction." The regression of Y on X means the prediction of Y by X. **Linear regression** is a method for predicting a criterion variable from one or more predictor variables. In simple regression, the criterion is predicted from a single independent variable (called predictor variable in regression analysis) and the best-fitting straight line is of the form \( Y' = bX + A \) where \( Y' \) is the predicted score, \( X \) is the predictor variable, \( b \) is the slope, and \( A \) is the Y intercept. Typically, the criterion for the "best fitting" line is the line for which the sum of the squared errors of prediction is minimized. In multiple regression, the criterion is predicted from two or more predictor variables. A **regression coefficient** is the slope of the slope of the regression line in simple regression or the partial slope in multiple regression. **Multiple regression** is linear regression in which two or more independent (predictor) variables are used to predict the dependent variable (called ‘criterion’ in regression analysis).

**Relative benefit increase (RBI)** is the increase in probability of a beneficial outcome (i.e., \( RBI = (EER - CER)/CER \) or \( RBI = BR - 1 \)). RBI can inflate the appearance of intervention effectiveness as it doesn’t reflect the baseline benefit of the outcome in the population. The RBI should be accompanied by a 95% confidence interval (CI).

**Relative frequencies** – see Conditional probabilities

**Relative risk (RR)** / **Relative benefit (BR)** (RR is also called Risk ratio): the ratio of the probability of the event occurring in the intervention group compared to the probability of the event occurring in the control group. (i.e., RR or BR = EER/CER). RR=1 or BR=1 means there is no difference between the two groups. RR or BR less than 1 indicates lower risk / benefit due to the intervention. RR or BR more than 1 means the intervention increased the risk / benefit. Sometimes RR that are decimals are expressed as a percentage; e.g. where RR=0.1 or 10%.

**Relative risk reduction (RRR)** is the reduction in risk of a negative outcome (i.e., \( RRR = (CER - EER)/CER \) or \( RRR = 1 - RR \)). RRR can inflate the appearance of intervention effectiveness as it doesn’t reflect the baseline risk of the outcome in the population. The RRR should be accompanied by a 95% confidence interval (CI).

**Reliability** in the broad scientific sense, refers to the consistency and trustworthiness of research findings in addition to the reproducibility of research procedures at other times (Hammersley, 1992; Kvale & Flick, 2007). Reliability in quantitative research is the extent to which a measurement instrument is dependable, stable and consistent when repeated under identical conditions.

**Reporting bias** or **Publication bias** is a type of bias created by the tendency to publish research reports that are statistically significant or otherwise support the preferred result. This form of bias also occurs when published studies present biased results, such as studies that only include such as only including outcomes where statistically significant differences were found. Failure to include unpublished studies in systematic reviews can bias results e.g., the true effect of an intervention may be overestimated. Publication bias may be represented graphically using a funnel plot or statistical methods.

**Representative sample** refers to the sample chosen to match the qualities of the population from which it is drawn. With a large sample size, random sampling will approximate a representative sample; stratified random sampling can be used to make a small sample more representative.

**Research aim/s** are concise statement/s that clearly outline the main purpose of the study (Thomas & Hodges, 2010).
**Research design** (also called ‘study design’) refers to the overall procedural plan adopted by the researcher to answer the research question validly, accurately and economically. The design helps to determine the proposed research path of the study including sampling, data collection and analysis.

**Research** in health is a planned, systematic process that generates new knowledge and refines and/or validates existing knowledge in such a way that health problems, practices and/or outcomes are influenced either directly or indirectly (Minichiello, Sullivan, Greenwood, & Axford, 2004).

**Research problem/s** concisely describe a specific issue, situation or knowledge/practice gap that need to be addressed in some potentially beneficial way

**Research process** is the set of step-by-step procedures that guide research.

**Research proposal**, the overall plan for the conduct of a project; before commencing a project the proposal usually needs to be approved in some way (e.g. by a research supervisor, ethics committee/s and/or funding agency).

**Research questions** are written as questions with ending in a question mark (?). They specify the key terms/variables in a way that indicates the focus of the study; should be written in a structured format such as ‘PICO’ or variations (e.g. PICOT). See **Structured questions**.

**Response bias** is a form of **Selection bias** which arises from refusal or non-response ranging between study groups. If the rate of response is also related to exposure status, then bias could be a reasonable explanation for any observed association. For example, if response to a questionnaire is lower for lower socio-economic groups and socio-economic status is a risk factor for the outcome under study (as it usually is) selection bias may be a problem when interpreting results.

**Restriction** is a way of controlling the effect of **confounding**; it limits entrance into the study of subjects who fall within a specified category of confounders. The disadvantages of this method are that it reduces the pool of eligible subjects and it does not permit later evaluation of the confounding effect. Since age is a confounder for many conditions (chronic disease, for example), ensuring that both exposure and comparison groups have the same age distribution is a way of using restriction to control for age as a confounder.

**Retrospective – prospective designs** look backwards and then forwards in time, i.e. data from the past is collected from pre-existing records or participant recall, then the treatment/intervention is introduced and the impact of that introduction is assessed from data gathered from the same participants in the future

**Retrospective cohort study** is where cohorts (i.e., groups of people with the presence or absence of an exposure or **risk factor**) are defined at a point of time in the past and information collected on subsequent outcomes usually from files or databases. E.g., the use of medical records to identify a group of women using oral contraceptives five years ago, and a group of women not using oral contraceptives, and then contacting these women or identifying in subsequent medical records the development of deep vein thrombosis.

**Retrospective designs** look backwards in time, i.e. the problem/situation being investigated is in the past. Data is collected either from a past time period or the data is based on participant’s recall of the problem/situation

**Review bias** occurs when the person interpreting a diagnostic test is aware of the results of the other test/s (i.e. the interpreter is not blinded). Review bias is common where test interpretation is
more subjective (i.e. physical examination); bias may also occur when the interpreter has knowledge of other previously done tests. **Diagnostic studies** may be unblinded, single blinded (one or other test is blinded) or double blinded. Lack of blinding tends to cause overestimation of **index test** accuracy. See **Blind, independent comparison**

**Rigor/rigour** in research refers to whether the researcher has actively carried out the study to the highest possible standard (Rees, 2011). Rigor requires careful adherence to the individual characteristics of the chosen methodology including ethical standards in such a way the research can be considered trustworthy (valid) and reliable (Liamputtong, 2017). It also includes recognising weaknesses in research design and reducing their impact as much as possible (Rees, 2011). The result of rigorous research should be an accurate and professional research report.

**Risk factor/s** are exposures, characteristics or attributes (e.g. high blood pressure, smoking, obesity) that have an association with the likelihood of developing a disease or infection. Risk factors are not necessarily the same as **prognostic factors** which indicate a better/worse outcome (prognosis) of a disease.

**Sample** is the total number of study participants/patients chosen from the overall population. Note, in quantitative research the study participants are often called ‘subjects’.

**Sample size**, see **Statistical Power**, **Statistical significance**, **Radom error**, **Representative sample** and **Power**

**Sampling error** is the likelihood that one sample is not totally representative of the population from which it came.

**Sampling** is the way that the subset of the population is selected.

**Scientific method**, a primarily deductive approach to logic tests theories initially posed through inductive logic (which begin from an initial observation). It is characterised by: 1) observation of phenomena; 2) invention of hypothesis/description consistent with observation; 3) use of hypothesis to predict; 4) testing predictions by experiment/observation; 5) adjustment of hypothesis; 6) repetition of 4) and 5) until no discrepancies between prediction and observation. The scientific method reflects positivist assumptions that science progresses by reducing from less basic to more basic and that it is value-free and objective. See **Positivist paradigm**

**Search strategy**, this is how you communicate to with your chosen database/s; it includes the terms you want to look for and how you want to look for them including a single **search ‘string’** and **delimiters**. Your search strategy acts as a record of the search process that you can use to find ‘the current best available evidence’; this record should also include names of the databases searched.

**Search ‘string’**: includes all variants of the search terms with **wildcards** and **truncations** that are grouped using **bracketing** and connected by **Boolean operators**. For example, (Female* OR Wom*n) AND (Brief intervention* OR Brief advice OR Brief counsel*) AND (Smoking cessation OR Stop* smoking OR Quit* smoking).

**Secondary data** is data that has already been collected through some other research process and is being used by a different researcher. Databases from hospitals and the data from censuses are examples of secondary data.

**Secondary evidence** is a synthesis, summary and/or critical appraisal of primary (original) research or topics; this evidence often includes recommendations for practice. See **Systematic reviews**.
Selection bias is a systematic error in participant recruitment and/or group allocation that causes baseline characteristic differences between the groups; i.e., it occurs when there is a systematic difference between the characteristics of the people selected for a study and the characteristics of those who are not. Relevant in observational cohort studies (e.g. prognostic studies) where selection is not randomised; bias in group allocation/recruitment causes differences in prognosis between the groups. In general, selection bias is a bigger problem for retrospective studies since both exposure and outcome have occurred at the time selection occurs for the study. See Spectrum bias, Response bias

Sensitivity measures how effective the test is in detecting the disease in those who truly have the disease. It is the probability that the test gives a positive result in people with the disease; it is calculated by dividing all the true positives (TP) by the true positives (TP) plus the false negatives (FN), i.e. TP/(TP+FN). If the cut-off for having the disease is changed to make it easier to detect the disease, the sensitivity will raise because more disease is detected.

Single subject experimental designs are non-observational, longitudinal analytic study designs with only one participant; they focus solely on that individual’s response to treatment over time (Liamputtong, 2017).

Size of Intervention effect is the amount of difference that an intervention makes to an outcome. It is reported differently depending on whether outcomes variables are continuous or dichotomous.

Smallest worthwhile effect is the smallest intervention effect that is clinically worthwhile. It is established by the health professional in differing ways often by referring to patient/s and relevant others. It is used along with the confidence intervals (CI) of the intervention effect to determine the clinical significance.

SnNOut is a mnemonic that stands for Sensitivity-Negative-Out; it means that when a diagnostic test has high sensitivity and its post-test probability is negative, the disease is most likely (but not always) going to be ruled ‘out’ i.e. truly not present.

Specificity measures how effective the test is in diagnosing that the disease is not present in those without the disease. It is the probability that the test gives a negative result in people without the disease; it is calculated by dividing all the true negatives (TN) by the true negatives (TN) plus the false positives (FP), i.e. TN/(TN+FP). If the cut-off for having the disease is changed to make it easier to detect the disease, the specificity will decrease because there will be more error.

Spectrum bias is a form of selection bias that can change the sensitivity and/or the specificity of the index test; it occurs when the study population is very different from the clinically relevant population in which the index test will be used. Such a difference may occur when, for example, recruitment is through a tertiary hospital with high rates of the disease but the test is to be used by a clinic-based population. Another example is when case-control designs are used where extreme contrast exists between the two groups (i.e. one group severely diseased and the other group very fit). Extreme contrast between groups in case-control studies can cause overestimation of accuracy.

SpPin is a mnemonic that stands for Specificity-Positive-In; it means that when a diagnostic test has high specificity and its post-test probability is positive, the disease is most likely (but not always) going to be ruled ‘in’ i.e. truly present. NOTE: When prevalence is low, as it often is in screening programs, then the positive predictive value will be low (i.e. most positive results will be false positives) regardless of whether the sensitivity and specificity are high.
Standard deviation (SD) is a widely-used measure of variability that shows how dispersed a set of data is from the mean; a higher SD means that the data is more spread apart. The SD is usually expressed with mean (e.g., a mean of 8 with SD = 2 is a range in results of 6 to 10). An important attribute of the standard deviation is that if the mean and standard deviation of a normal distribution are known, it is possible to compute the percentile rank associated with any given score. Properties of the normal distribution curve predict that approximately 95% of the population will record a value of [mean ± 2 standard deviations].

Statistical heterogeneity in a systematic review refers to differences in intervention effects that arise because of clinical and/or methodological differences between individual studies (i.e. clinical and/or methodological heterogeneity). Although some variation in the effects of interventions between studies will always exist, whether this variation is greater than what is expected by chance alone needs to be determined. See Heterogeneous

Statistical Power is the probability of obtaining a p value less than 0.05 (p < 0.05) with a given sample size and a given effect size (e.g. strength of correlation within the "population" of interest). It involves calculating the probability of NOT committing a Type II error (β value = failure to find and report a relationship between variables when a relationship actually does exist). At a given significance level, the power of the test is increased by having a larger sample size. The minimum accepted level is 80%, meaning that there a chance of eight in ten that a difference of the specified effect size will be detected.

Statistical significance is the extent to which the results of a study could have happened by random error rather than as the result of an intervention or independent variable. A statistically significant result means that results are unlikely to have occurred by chance, so there is a likely real difference between the groups; this implies that the hypothesis is true (and the null hypothesis is not true). Put another way statistical significance is a measurement of the likelihood of a chance uneven distribution of known and unknown confounder variables (i.e. Random error) across experimental groups relative the groups’ sample size. The difference between groups is measured in two ways. 1) How probable (p) that the result is due to chance, indicated by the p value (usually p<0.05). 2) How confident we can be that true value of the results are within a certain range of values, this is the confidence interval (CI); statistical significance is usually 95% CI, CI do not include the ‘no effect value’.

Stratified random sampling involves dividing the population into two or more subgroups (or strata). Random samples are then taken from each subgroup with sample sizes proportional to the size of the subgroup in the population. For instance, if a population contained equal numbers of men and women, and the variable of interest is suspected to vary by gender, one might conduct stratified random sampling to ensure a representative sample.

Structured abstracts are summarised descriptions of published articles where the details are arranged systematically into paragraphs with headings that address key aspects of the study, often journals dictate what those headings are. The IMRAD format of four headings is commonly used as guide: Introduction, Methods, Results, Discussion.

Structured questions are directly relevant to an information need and take a format designed to guide the search for evidence that answers the information need. These questions should sound like
a question and end with a question mark (?). Possible formats include the ‘PICO’ format and variations. See Research questions

Study design (also called ‘research design’), refers to the procedural plan adopted by the researcher to answer questions validly, accurately and economically. The design helps to determine the proposed research path of the study including sampling, data collection and analysis.

Subgroup analyses are analytic observational (i.e. nonrandomised); they involve splitting of participant data into relevant subgroups (e.g., subsets of participants or subsets of studies), usually to make comparisons between the groups. Used to investigate heterogeneity and to answer specific questions about individual groups (e.g., participants, types of intervention, types of study). Characteristics of subgroups should be identified prior to commencement of study.

Systematic Reviews are a form of secondary evidence that involves the systematic location, appraisal and synthesis of all relevant primary research evidence. They focus on a clinical topic and answer a specific research question; the methodology used must be reproducible. An extensive, systematic literature search is conducted to identify published and un-published studies with sound methodology that meet a prearranged set of criteria. The studies are reviewed, assessed for quality including risk of bias, and the results summarized in line with a predetermined protocol and the specific research review question. A meta-analysis of quantitative data or a meta-synthesis of qualitative data may be undertaken. (Higgins & Green, 2011). Occasionally, ‘reviews of reviews’ or ‘umbrella reviews’ will be undertaken; these are systematic reviews of other systematic reviews (The Joanna Briggs Institute, 2014).

Thematic analysis is a systematic process for identifying patterns or themes within the data.

Theoretical sampling involves choosing the sample based on the participant characteristics that are most relevant to the emerging theory (Minichiello et al., 2004). Theoretical sampling is a non-probability sampling method used in qualitative research, particularly during grounded theorising; this is an iterative process requiring near simultaneous data collection, analysis and theorising.

Theories are orderly ways of structuring knowledge so as to describe, explain and/or predict phenomena, events and activities (Bryar & Sinclair, 2011; Scott & Marshall, 2005). Theories ideally comprise a creative and rigorous structuring of ideas (concepts) that are inter-related via propositions. A theory may be deduced using existing knowledge from other disciplines or inductively developed through reflective practice and/or research.

Thesis statement is a statement made to introduce the main idea or theory being argued in an essay, report or research proposal. Each part of the essay, report or proposal should discuss and move toward support of the main thesis statement. The thesis statement is also provided in the conclusion. A doctoral thesis or masters thesis is the long report used to argue a particular thesis statement.

Transferability in qualitative research considers how well the findings can be applied (or generalised) to other groups or specific settings (Liamputtong, 2013).

Truncations allow inclusion of variant spellings or different word endings e.g., Neoplas* will find neoplasia, neoplastic, neoplasm

Trustworthy research see Validity, Reliability and Rigor/rigour

t-test: Statistical test that compares two groups where the dependent variable is interval data and conforms to the normal distribution (i.e. it is a parametric test). There are two types: the
independent (unrelated) t-test compares the same measurements taken from two separate groups; the paired (related) t-test compares the same measurements from the same group at different points in time, or from matched individuals.

**Type I error** (also referred to as the \( \alpha \) value) is when a relationship between variables (or groups or diagnostic tests) is found and reported when a relationship does not actually exist.

**Type II error** (also referred to as the \( \beta \) value) is the failure to find and report a relationship between variables (or groups or diagnostic tests or **prognostic factors**) when a relationship does exist.

**Univariate analysis** is the examination of a single variable at a time across all cases. The aim is to describe the study sample using a few easy to understand numbers to gain a basic picture of the sample. The key processes are a) **distribution**, b) **central tendency** and c) **dispersion**.

**Validity** in the broad scientific sense, refers to the extent to which a concept, measure, outcome or knowledge claim is well founded and corresponds accurately to ‘truth’ or ‘the real world’ (Hammersley, 1990; Maxwell, 2013; Porter, 2007). Validity in quantitative research refers to the degree to which the scale or instrument measures what it purports to measure (Liamputtong, 2017).

**Variability** is the extent to which the values for the variables differ from one another. That is, how much they vary. Variability can also be thought of as how spread out a distribution is. The standard deviation and the semi-interquartile range are measures of variability.

**Variables** are measurable versions of **concepts**; they are created by writing ‘operational definitions’. Variables, as opposed to concepts, respond to variations within the real world and take on different values according to the operational definition (Watt & van den Berg, 2002). In quantitative research these different values are numeric (or converted to numbers) whereas in qualitative research the different values are only expressed using words (or some non-numeric media).

**Variance** is a widely used statistical measure of **variability**. It is defined as the mean squared deviation of scores from the mean so it indicates the spread of data (i.e. how close each item of data is to the mean data point); it is calculated by calculating the average of squared differences from the mean. Data that is close to the mean is indicated by a small variance and data that is widely spread from the mean has a large variance.

**Verification bias** (also called ‘workup bias’ or ‘referral bias’) occurs when the **index test** is not followed by the **reference standard** test in all participants. Sometimes only one or other test is given; often, if the index test is negative the ‘gold standard’ test is omitted for pragmatic reasons. Verification bias tends to raise **sensitivity** and lower **specificity**; it can be statistically adjusted for if all participant numbers are known including who had and who did not have each type of test. In retrospective studies, sometimes patients are deliberately excluded from participation based on only having one of the tests.

**Wildcards** allow substitution of one character in the word e.g., *edema will find the word oedema as well as the word edema; wom*n will find women and woman; p*ediatrics will find the word pediatrics as well as the word paediatrics.
References:


Merlin, T., Weston, A., & Tooher, R. (2009). Extending an evidence hierarchy to include topics other than treatment: revising the Australian 'levels of evidence'. *BMC Medical Research Methodology, 9*(1), 34. doi: [http://www.biomedcentral.com/1471-2288/9/34](http://www.biomedcentral.com/1471-2288/9/34)


Payne, A. (n.d.). *Evidence Based Practice or Practice Based Evidence what is the difference?*: Thematic Network for Dietetics.


