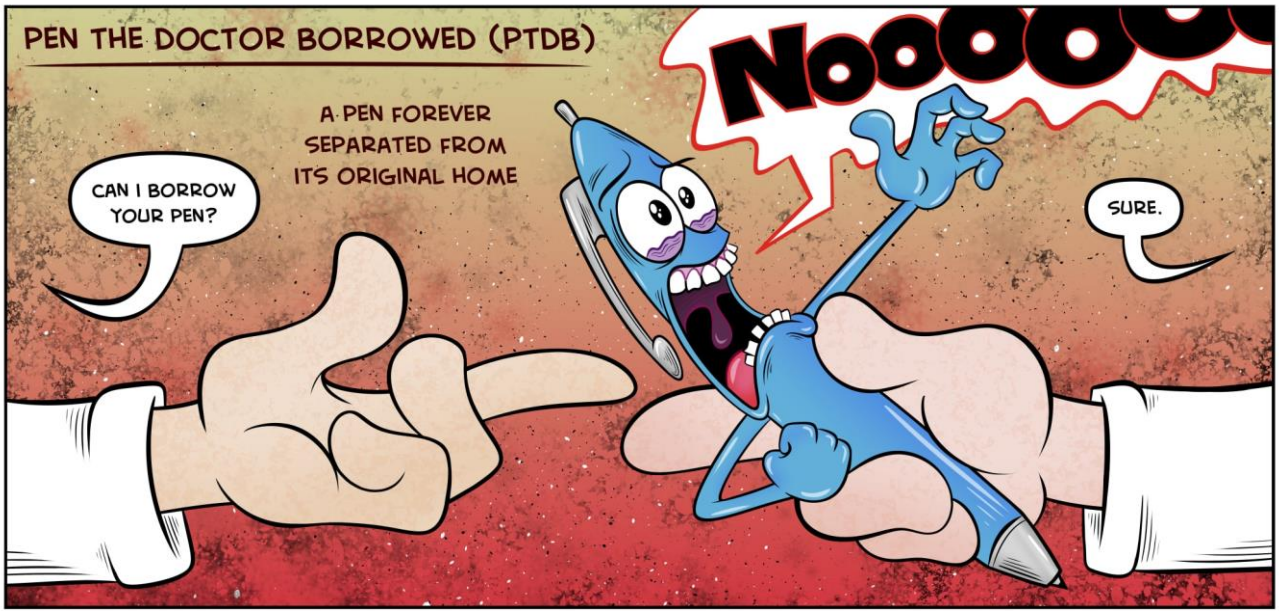


Audit and Research

Clinical audit

Quality improvement process that seeks to improve patient care and outcomes through systematic review of care against explicit criteria and the implementation of change. Aspects of the structure, processes, and outcomes of care are selected and systematically evaluated against explicit criteria. Where indicated, changes are implemented at an individual, team, or service level and further monitoring is used to confirm improvement in healthcare delivery. (NICE).

Research

Aims to derive new knowledge which is potentially generalisable or transferable.

6 pillars of clinical governance:

- Clinical effectiveness
- Research and development
- Openness
- Risk management
- Education and training
- Clinical audit

Audit Categories

Audits may be used in a variety of clinical settings. These range from standards based audits, which will be familiar to most clinicians, through to systems based audits which focus more on the processes within an organisation.

Types of audit

Financial audit	A historically oriented, independent evaluation performed for the purpose of attesting to the fairness, accuracy, and reliability of financial data
Operational audit	A future-oriented, systematic, and independent evaluation of organizational activities. Financial data may be used, but the primary sources of evidence are the operational policies and achievements related to organizational objectives. Internal controls and efficiencies may be evaluated during this type of review.
Departmental review	A current period analysis of administrative functions, to evaluate the adequacy of controls, safeguarding of assets, efficient use of resources, compliance with related laws, regulations and institutional policy and integrity of financial information.
Standards based audit	Comparison of care or passage of care against set and widely agreed standards or outcomes.
Systems based audit	Evaluation of processes occurring within an institution.

Systems based audits are an integral part of the process of clinical governance.

Audit vs Research

Clinical audit is 'a quality improvement process that seeks to improve patient care and outcomes through systematic review of care against explicit criteria and the implementation of change'.

Research is a one-off, systematic and organised way to find answers to questions. Research does not check whether you are complying with standards, instead its aim is to create new knowledge and new standards.

In essence research helps to establish best practice while audit ensures that best practice is carried out.

Consent

There are 3 types of consent: (Informed, Expressed, Implied)

Consent forms used in UK NHS

Consent Form 1	For competent **adults** who are able to consent for themselves where consciousness may be impaired (e.g. GA)
Consent Form 2	For an adult consenting on behalf of a **child** where consciousness is impaired
Consent Form 3	For an adult or child where **consciousness is not impaired**
Consent Form 4	For adults who **lack capacity** to provide informed consent

Capacity

Key points include:

1. Understand and retain information
2. Patient believes the information to be true
3. Patient is able to weigh the information to make a decision

All patients must be assumed to have capacity

Consent in minors

Young children and older children who are not Gillick competent cannot consent for themselves. In British law the patients biological mother can always provide consent. The child's father can consent if the parents are married (and the father is the biological father), or if the father is named on the birth certificate **(irrespective of marital status)**.
If parents are not married and the father is not named on the birth certificate then the father cannot consent.

NB

If an adult patient has capacity, you cannot overrule their decision to refuse treatment even if that will lead to their death or a severe permanent injury.

Proceed without consent *only if the patient needs emergency treatment to save their life,* ***but they're incapacitated*** *(for example, they're unconscious) – the reasons why treatment was necessary should be fully explained once they have recovered. A consent type 4 is usually done for unconscious patients that need lifesaving treatment.*

Gillick competency used to decide whether a child (under 16 years of age) is able to consent to their own medical treatment, without the need for parental permission or knowledge.
i.e. ***Children under the age of 16 can consent to their own treatment*** *if they're believed to have enough intelligence, competence and understanding to fully appreciate what's involved in their treatment. This is known as being Gillick competent.*

The Bolam test is a test that can be carried out to ascertain whether a doctor or other medical professional has breached their duty of care to a patient. All medical professionals have a duty of care towards patients in so much as they must do what they can to keep them safe from harm.

Court decision / Refer to court

For any unclear or controversial cases, referral to a court maybe necessary. For example

- Deciding to stop life support / treatment for patients who haven't made an advance decision.
- If parents refuse a treatment that is in the child's best interest
- If a young person who is Gillick competent refuses

Further reading

- *https://www.nhs.uk/common-health-questions/nhs-services-and-treatments/do-i-have-the-right-to-refuse-treatment/*
- *https://www.nhs.uk/conditions/consent-to-treatment/children/*
- *https://www.nhs.uk/conditions/consent-to-treatment/*
- *https://www.nhs.uk/conditions/end-of-life-care/advance-decision-to-refuse-treatment/*

Cluster Randomised Controlled Trials

- Groups are randomised rather than individuals
- Avoids cross contamination amongst participants
- Participants in any one cluster are more likely to respond in a similar fashion
- **Higher risk of unit of analysis error** as these studies should be analysed as clusters rather than on an individual basis. This leads to a higher false positive rate.
- It is possible to adjust for clustering in statistical analyses

Incidence and Prevalence

These two terms are used to describe the frequency of a condition in a population.

The **incidence** is the number of **new** cases per population in a **given time period**.

For example, if condition X has caused 40 new cases over the past 12 months per 1,000 of the population the annual incidence is 0.04 or 4%.

The **prevalence** is the **total** number of cases per population at a **particular point in time**.

For example, imagine a questionnaire is sent to 2,500 adults asking them how much they weigh. If from this sample population, 500 of the adults were obese then the prevalence of obesity would be 0.2 or 20%.

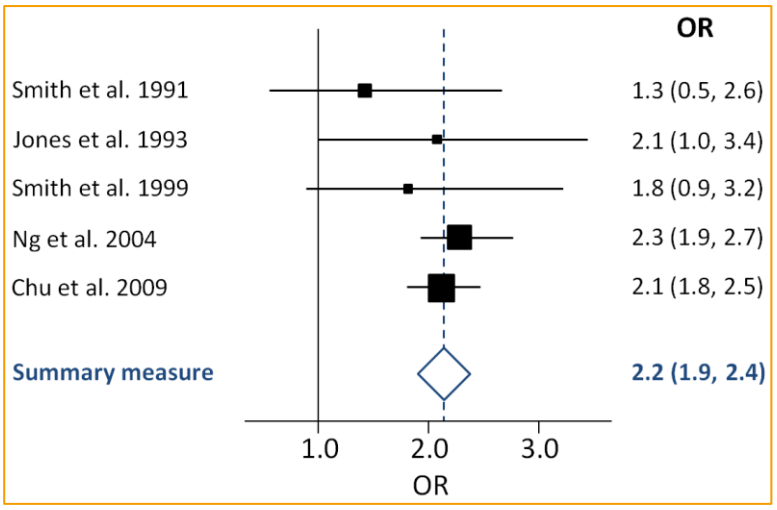
Relationship

- Prevalence = incidence * duration of condition
- In chronic diseases the prevalence is much greater than the incidence
- In acute diseases the prevalence and incidence are similar. For conditions such as the common cold the incidence may be greater than the prevalence

Forest Plots

A Forest plot is a graphical display designed to illustrate the relative strength of treatment effects in multiple quantitative scientific studies, addressing the same question. It is often used to graphically **display meta analyses of RCTs**.

The graph may be plotted on a natural logarithmic scale when using odds ratios or other ratio-based effect measures, so that the confidence intervals are symmetrical about the means from each study and to ensure undue emphasis is not given to odds ratios greater than 1 when compared to those less than 1. The area of each square is proportional to the study's weight in the meta-analysis. The overall meta-analysed measure of effect is often represented on the plot as a vertical line. This meta-analysed measure of effect is commonly plotted as a diamond, the lateral points of which indicate confidence intervals for this estimate.

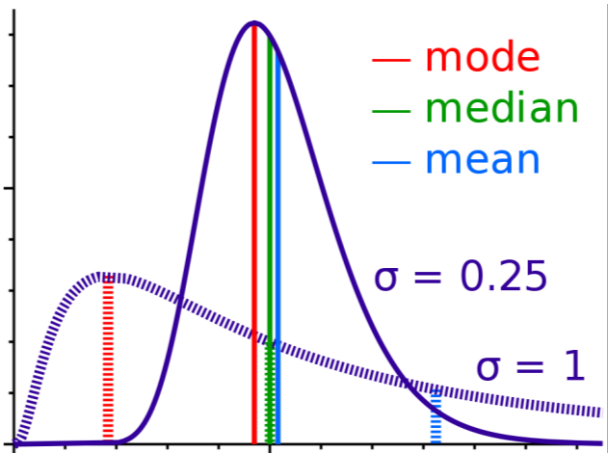


Generic Forest plot

A vertical line representing no effect is also plotted. If the confidence intervals for individual studies overlap with this line, it demonstrates that at the given level of confidence their effect sizes do not differ from no effect for the individual study. The same applies for the meta-analysed measure of effect: if the points of the diamond overlap the line of no effect the overall meta-analysed result cannot be said to differ from no effect at the given level of confidence.

Normal Distribution

The normal distribution is also known as the Gaussian distribution or 'bell-shaped' distribution. It describes the spread of many biological and clinical measurements

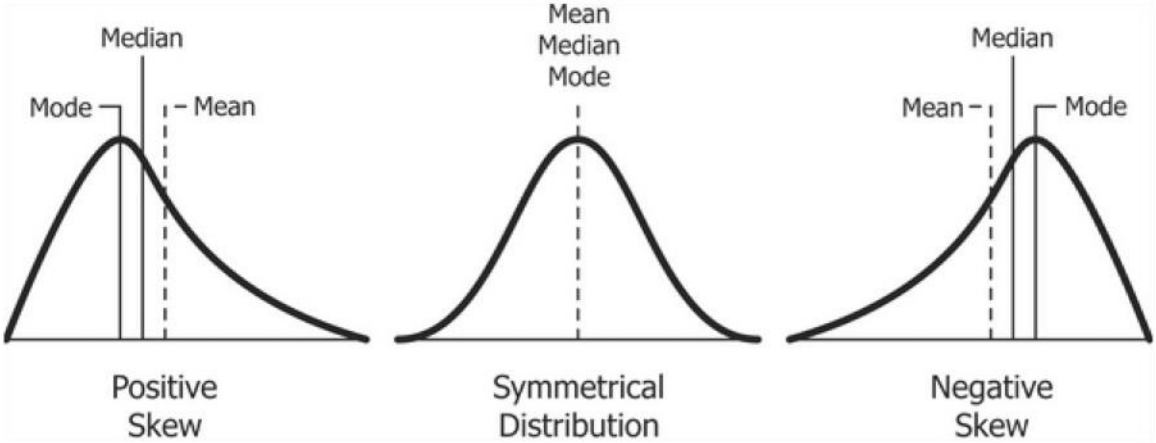


Properties of the Normal distribution

- symmetrical i.e. Mean = mode = median
- 68.3% of values lie within 1 SD of the mean
- 95.4% of values lie within 2 SD of the mean
- 99.7% of values lie within 3 SD of the mean
- This is often reversed, so that within 1.96 SD of the mean lie 95% of the sample values
- The range of the mean - (1.96 *SD) to the mean + (1.96 * SD) is called the 95% confidence interval, i.e. If a repeat sample of 100 observations are taken from the same group 95 of them would be expected to lie in that range

Standard deviation

- the standard deviation (SD) represents the average difference each observation in a sample lies from the sample mean
- SD = square root (variance)



Mode = most frequent, *e.g. 13, 18, 13, 14, 13, 16, 14, 21, 13 (Mode = 13)*
Median = middle value, *e.g. 13, 18, 13, **14**, 13, 16, 14, 21, 13 (Median = 14)*
Mean = average value, *e.g. 13, 18, 13, 14, 13, 16, 14, 21, 13 (Mean =135 /9 = 15)*

Pre and Post Test Odds and Probability

Pre-test probability

The proportion of people with the target disorder in the population at risk at a **specific time** (**point prevalence**) or time interval (**period prevalence**)

For example, the prevalence of rheumatoid arthritis in the UK is 1%

Post-test probability

The proportion of patients with that particular test result who have the target disorder
Post-test probability = post test odds / (1 + post-test odds)

Pre-test odds

The odds that the patient has the target disorder before the test is carried out
Pre-test odds = pre-test probability / (1 - pre-test probability)

Post-test odds

The odds that the patient has the target disorder after the test is carried out
Post-test odds = pre-test odds x likelihood ratio
where the likelihood ratio for a positive test result = sensitivity / (1 - specificity)

Qualitative and Quantitative Data

Qualitative and quantitative data

Qualitative (categorical) data refers to different descriptions of a characteristic, although it may be possible to allocate a number it has no scale.

Quantitative data is associated with numerical values on a numerical scale.

Since quantitative data is based on a numerical scale it can be organised to create a distribution curve. The central tendency may be estimated using the mode, median and mean. The standard deviation gives an estimation of the spread of data.

Relative Risk

Relative risk (RR) is the ratio of risk in the experimental group (experimental event rate, EER) to risk in the control group (control event rate, CER)

$$RR = \frac{EER}{CER}$$

To recap

- EER = rate at which events occur in the experimental group
- CER = rate at which events occur in the control group

For example, if we look at a trial comparing the use of paracetamol for back pain compared to placebo we may get the following results

	Total number of patients	Experienced significant pain relief
Paracetamol	100	60
Placebo	80	20

Experimental event rate, EER = 60 / 100 = 0.6
Control event rate, CER = 20 / 80 = 0.25
Therefore the relative risk = EER / CER = 0.6 / 0.25 = 2.4

$$Absolute\ Risk\ Change = EER - CER$$

If the risk ratio is > 1 then the rate of an event (in this case experiencing significant pain relief) is increased compared to controls. It is therefore appropriate to calculate the relative risk increase if necessary (see below).

If the risk ratio is < 1 then the rate of an event is decreased compared to controls. The relative risk reduction should therefore be calculated (see below).

Relative risk reduction (RRR) or **relative risk increase (RRI)**
Is calculated by dividing the absolute risk change by the control event rate
Using the above data, RRI = (EER - CER) / CER = (0.6 - 0.25) / 0.25 = 1.4 = 140%

$$RRI = \frac{EER - CER}{CER}$$

Absolute Risk Reduction

The absolute risk reduction is the decrease in risk of a given activity or treatment in relation to a control activity or treatment. It is the inverse of the number needed to treat.

The absolute risk reduction is usually calculated for two different treatments. For example, consider surgical resection (X) versus watchful waiting (Y) for prostate cancer. A defined end point, such as 5-year survival is required. If the probabilities pX and pY of this end point are known, then the absolute risk reduction is calculated (pX-pY).

The inverse / reciprocal of absolute risk reduction is the *Number Needed to Treat*. This is useful in determining the cost Vs benefit of many treatments.

Number needed to treat (NNT)
Definition: how many patients would need to receive a treatment to prevent one event. It is the absolute difference between two treatments.

$$NNT = \frac{1}{Absoulte\ Risk\ Reduction}$$

Positive Predictive Values

The ***positive predictive value (PPV)*** is the probability that an individual with a positive screening result has the disease. The ***sensitivity*** is the probability that an individual with the disease is screened positive and the specificity is the probability that an individual without the disease is screened negative.

Screening tests

- Sensitivity: proportion of true positives identified by a test
- Specificity: proportion of true negatives correctly identified by a test
- Positive predictive value: proportion of those who have a positive test who actually have the disease
- Negative predictive value: proportion of those who test negative who do not have the disease

Predictive values are dependent on the prevalence

- Likelihood ratio for a positive test result = sensitivity/(1-specificity)
- Likelihood ratio for a negative test result = (1-sensitivity)/specificity

Likelihood ratios are not prevalence dependent

Screening Test Statistics

It would be unusual for a medical exam not to feature a question based around screening test statistics. The available data should be used to construct a contingency table as below:

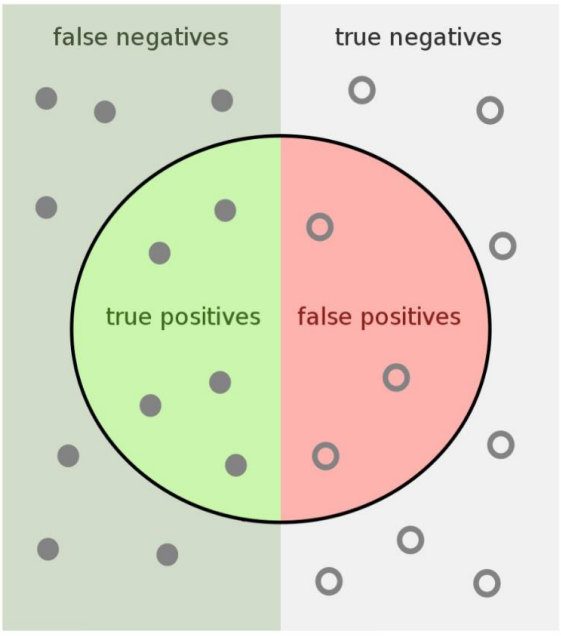
	Disease present	Disease absent
Test positive	TP	FP
Test negative	FN	TN

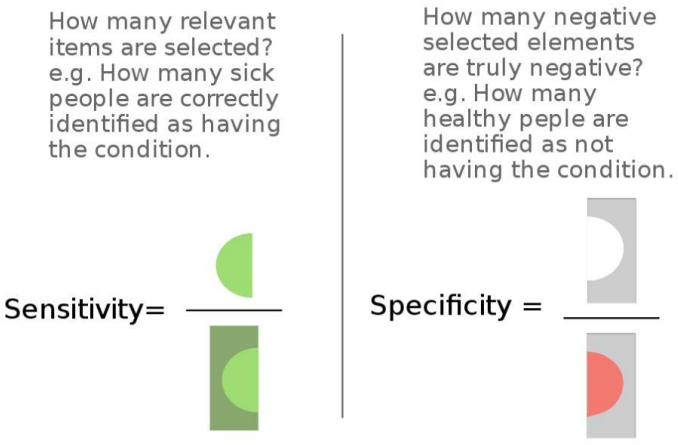
TP = true positive; FP = false positive; TN = true negative; FN = false negative

The table below lists the main statistical terms used in relation to screening tests:

Sensitivity	TP / (TP + FN)	Proportion of patients with the condition who have a positive test result
Specificity	TN / (TN + FP)	Proportion of patients without the condition who have a negative test result
Positive predictive value	TP / (TP + FP)	The chance that the patient has the condition if the diagnostic test is positive
Negative predictive value	TN / (TN + FN)	The chance that the patient does not have the condition if the diagnostic test is negative
Likelihood ratio for a positive test result	sensitivity / (1 - specificity)	How much the odds of the disease increase when a test is positive
Likelihood ratio for a negative test result	(1 - sensitivity) / specificity	How much the odds of the disease decrease when a test is negative

Positive and negative predictive values are prevalence dependent. Likelihood ratios are not prevalence dependent





Significance Tests

A null hypothesis (H_0) states that two treatments are equally effective (and is hence negatively phrased). A significance test uses the sample data to assess how likely the null hypothesis is to be correct.

For example:

- 'there is no difference in the prevalence of colorectal cancer in patients taking low-dose aspirin compared to those who are not'

The alternative hypothesis (H_1) is the opposite of the null hypothesis, i.e. There is a difference between the two treatments

The **p value** is the probability of obtaining a result by chance at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. It is therefore equal to the chance of making a type I error (see below).

Two types of errors may occur when testing the null hypothesis

- **Type I**: The null hypothesis is rejected when it is true - i.e. Showing a difference between two groups when it doesn't exist, a false positive. This is determined against a preset significance level (termed alpha). As the significance level is determined in advance the chance of making a type I error is not affected by sample size. It is however increased if the number of end-points are increased. For example, if a study has 20 end-points it is likely one of these will be reached, just by chance.
- **Type II**: The null hypothesis is accepted when it is false - i.e. Failing to spot a difference when one really exists, a false negative. The probability of making a type II error is termed beta. It is determined by both sample size and alpha.

	Study accepts H_0	Study rejects H_0
Reality H_0		Type 1 error (alpha)
Reality H_1	Type 2 error (beta)	Power (1 - beta)

The power of a study is the probability of (correctly) rejecting the null hypothesis when it is false

- power = 1 - the probability of a type II error
- power can be increased by increasing the sample size

Power Calculations and Statistical Error

Statistical error

Type 1 Error	• Test rejects true null hypothesis • Rate of type 1 error is the given the value of α • It usually equals the significance level of a test
Type 2 Error	• Test fails to reject a false null hypothesis • Rate of type 2 errors is given the value of β • It is related to the power of the test

Statistical power

The power of a test is the probability that the test will reject the null hypothesis when it is false (thereby avoiding a type 2 error). Increasing the power of a test will reduce the probability of a type 2 error. Usually a value of 0.8 is selected.

Type I Error (alpha)

Type II Error (beta)

Statistics & Statistical Tests

Statistics is a topic that generally strikes fear and dread into most surgeons' hearts. The MRCS is not an examination designed to test mathematical skill but the examiners do expect you to have working knowledge of commonly used tests so that you can appraise the literature properly.

Data types

Before selecting a method of statistical analysis it is imperative that the type of data to be analysed is correctly categorised. Commonly used terms include nominal, ordinal, interval and continuous.

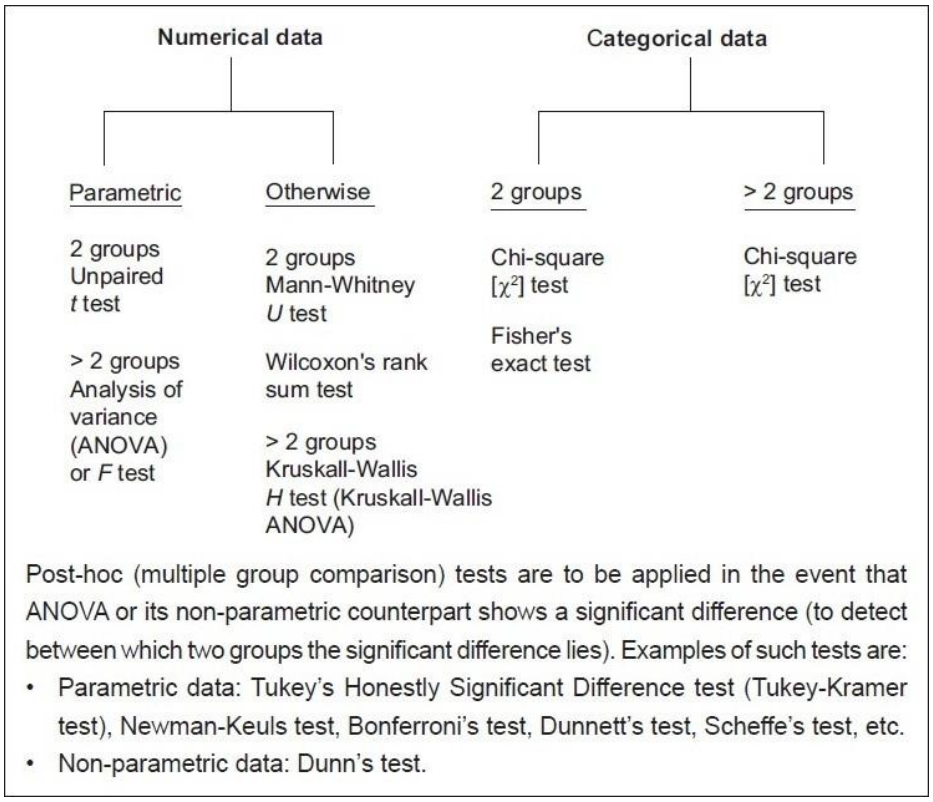
Term	Interpretation
Nominal	Data can be allocated a numerical code that is arbitrary. For example allocating people as alive or dead using codes of 0 or 1
Ordinal data	Data using numbers that can be used on a scale. Severity of pain is often measured in this way
Interval scale	Data is measured numerically. However, the zero point is arbitrary
Continuous	Data is measured numerically where the numerical value is a real number and may be any value. Examples include height and weight

Analysing data

Having ascribed the data it is then possible to begin the process of analysis. Nominal data is often tabulated into categories because of the nature of the underlying data sets. Continuous data may be displayed graphically often as individual data points. When the sample size is large enough, continuous data can be analysed to determine the distribution of the data points. Often, but not always these will be in the form of a gaussian distribution. Determining whether data is normally distributed or not is key to making sense of the subsequent statistical tests. Parametric tests are used to test normally distributed data, the **T Test** is one of the best examples. Data which is not normally distributed cannot be analysed in this way and a non-parametric test must be used. Examples of such tests include **Chi Squared and Mann Whitney U tests.** Chi squared tests often appear in the medical literature. There are some assumptions that are made in relation to Chi squared tests; these include the need to use 2 degrees of freedom (usually) and the minimum sample size. Where the sample size is small then a different test is appropriate and the **Fishers exact test** is often used. In situations where data is normally distributed and paired samples are taken from the same individuals (such as following an intervention) then the **paired T Test** may be used.

Multiple testing and post hoc analysis

In the ideal world statistical analysis is conducted on data that is collected prospectively according to preset power calculations and defined end points. Occasionally, data does not produce an expected outcome or a certain type of patient appears to have a different result. Subsequent analysis of such groups is termed a **post hoc analysis**. This can be perfectly legitimate, alternatively it can represent the last ditch attempt of a researcher to try and find any aspect of the data that is worthwhile. This can lead to errors and false rejection of a null hypothesis. A statistically significant result is more likely to occur if the same dataset is subjected to multiple analyses. To counteract this problem some researchers will apply a **Bonferroni correction**, this adjusts the analysis to allow for multiple testing.



Tests to address the question: Is there a difference between groups – unpaired (parallel and independent groups) situation?

Type of Test	Use
Correlational: these tests look for an association between variables	
Pearson Correlation	Tests for the strength of the association between two continuous variables
Spearman Correlation	Tests for the strength of the association between two ordinal variables (does not rely on the assumption of normally distributed data)
Chi-Square	Tests for the strength of the association between two categorical variables
Comparison of Means: these tests look for the difference between the means of variables	
Paired T-Test	Tests for the difference between two variables from the same population (e.g., a pre- and posttest score)
Independent T-Test	Tests for the difference between the same variable from different populations (e.g., comparing boys to girls)
ANOVA	Tests for the difference between group means after any other variance in the outcome variable is accounted for (e.g., controlling for sex, income, or age)
Regression: these tests assess if change in one variable predicts change in another variable	
Simple Regression	Tests how change in the predictor variable predicts the level of change in the outcome variable
Multiple Regression	Tests how changes in the combination of two or more predictor variables predict the level of change in the outcome variable
Non-Parametric: these tests are used when the data does not meet the assumptions required for parametric tests	
Wilcoxon Rank-Sum Test (Mann–Whitney U test)	Tests for the difference between two independent variables; takes into account magnitude and direction of difference
Wilcoxon Sign-Rank Test	Tests for the difference between two related variables; takes into account the magnitude and direction of difference
Sign Test	Tests if two related variables are different; ignores the magnitude of change—only takes into account direction

Paired vs Unpaired T-Test

Paired T-Test	Unpaired T-Test (Student's test)
Data is derived from study subjects who have been measured at two time points (so each individual has two measurements). The two measurements generally are before and after a treatment intervention e.g. measuring glucose concentration in diabetic patients before and after insulin injection	Applied to two independent groups e.g. diabetic patients versus non-diabetics In addition to the assumption that the data is from a normal distribution, there is also the assumption that the standard deviation (SD)s is approximately the same in both groups

Study Design

The following table highlights the main features of the main types of study:

Randomised controlled trial	Participants randomly allocated to intervention or control group (e.g. standard treatment or placebo) Practical or ethical problems may limit use
Cohort study	Observational and prospective. Two (or more) are selected according to their exposure to a particular agent (e.g. medicine, toxin) and followed up to see how many develop a disease or other outcome. The usual outcome measure is the **relative risk**. Examples include Framingham Heart Study
Case-control study	Observational and retrospective. Patients with a particular condition (cases) are identified and matched with controls. Data is then collected on past exposure to a possible causal agent for the condition. The usual outcome measure is the odds ratio. Inexpensive, produce quick results Useful for studying rare conditions Prone to confounding
Cross-sectional survey	Provide a 'snapshot', sometimes called prevalence studies Provide weak evidence of cause and effect

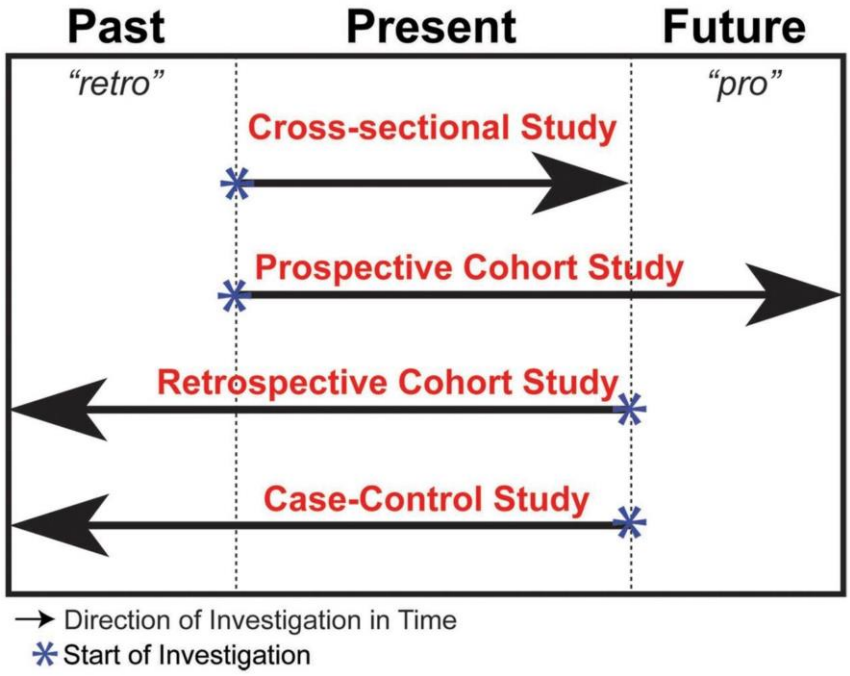
Data collection

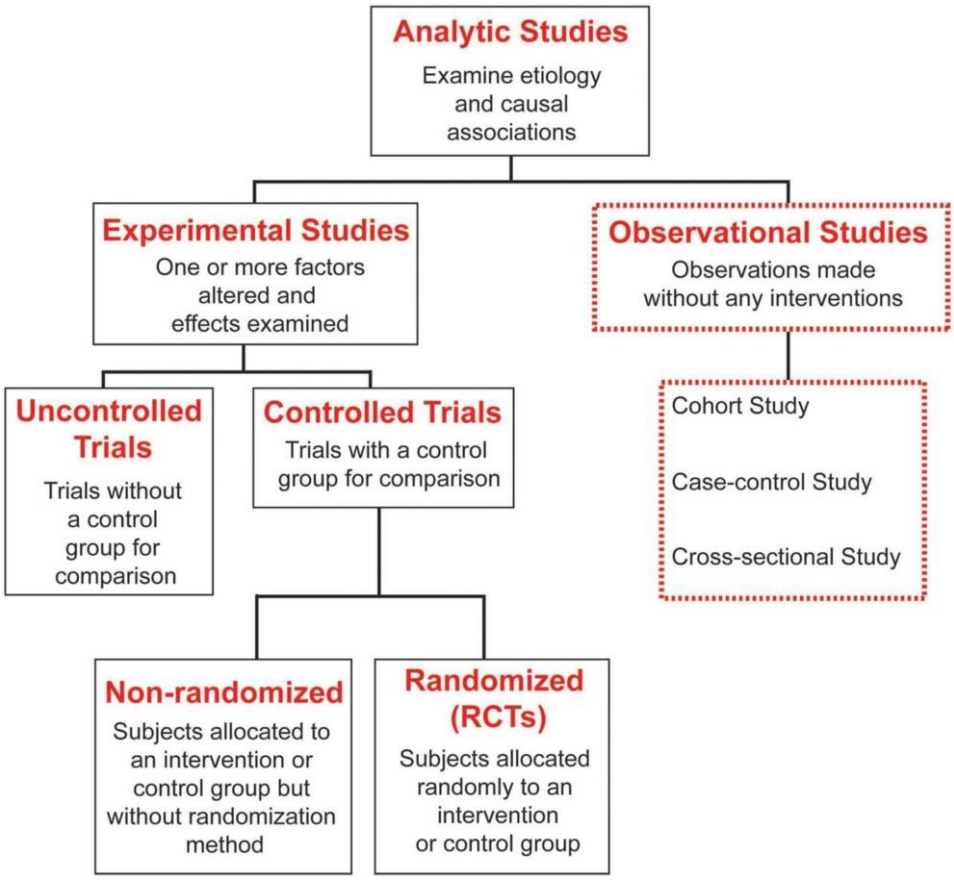
Prospective:

- Data collection starts with enrollment of patient in the study and for a specific time period

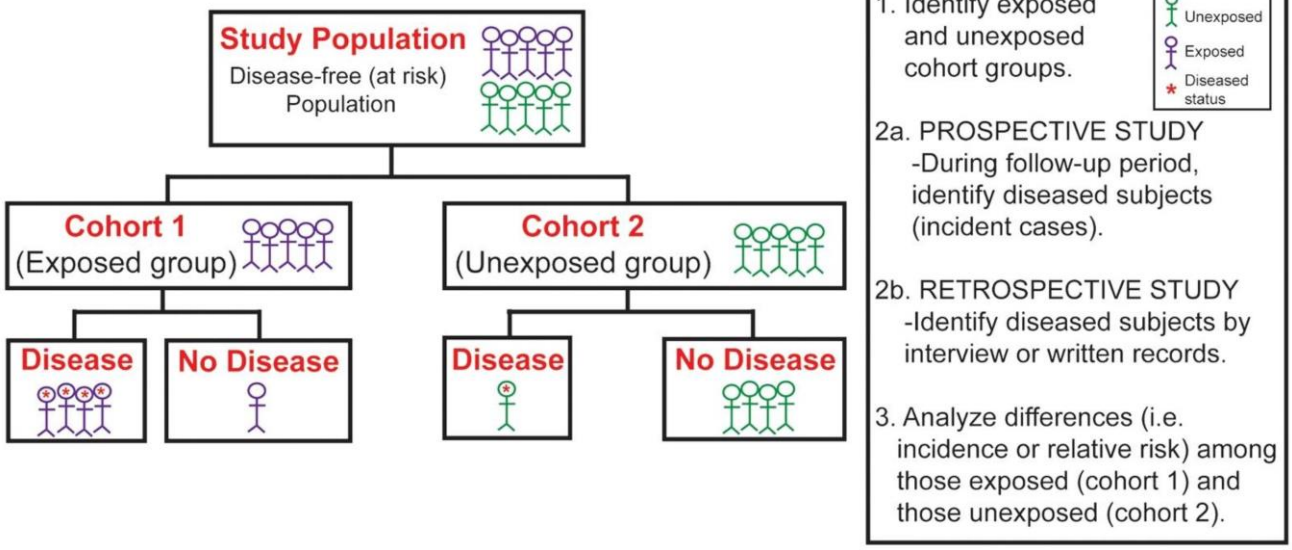
Retrospective:

- Patient is enrolled in the study after collection of data
- Data are taken from hospital records and therefore only those which are documented can be collected for the study

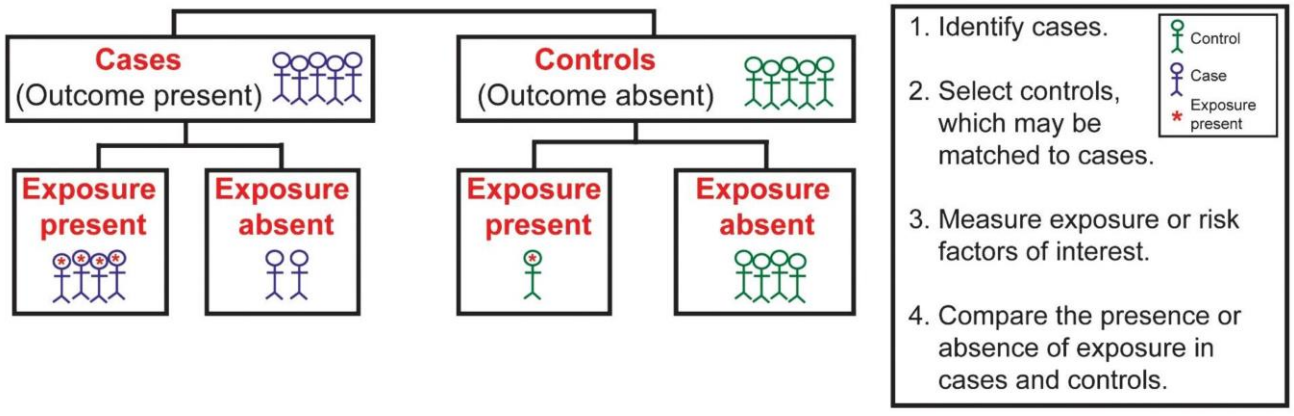




COHORT STUDY



CASE-CONTROL STUDY



Study Design: Evidence and Recommendations

Levels of evidence

- I - Evidence from meta-analysis of randomised controlled trials
- II - Evidence derived from at least one properly designed randomised controlled trial
- III - Evidence from correlation and comparative studies or use of historical controls
- IV - Evidence from case series or case reports
- V - Expert opinion or founded on basic principles

Knowledge of the sub groups of the levels of evidence are not routinely tested in MRCS Part A.

Grading of recommendation

- Grade **A** - based on evidence from at least one RCT (i.e. Ia or Ib)
- Grade **B** - based on evidence from non-RCTs (i.e. IIa, IIb or III)
- Grade **C** - based on evidence from a panel of experts (i.e. IV)