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) is a measure of how much dispersion exists from the mean
- SD = square root (variance)

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 of 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

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). The term relative risk ratio is sometimes used instead of relative risk.

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 dysmenorrhoea 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 ratio = EER / CER = 0.6 / 0.25 = 2.4

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%

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

Funnel plots : Demonstrate the existence of publication bias in meta-analyses

Standard error of the mean : standard deviation / square root (number of patients)

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:

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

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

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

Numbers needed to treat and absolute risk reduction

Numbers needed to treat (NNT) is a measure that indicates how many patients would require an intervention to reduce the expected number of outcomes by one

It is calculated by 1/(Absolute risk reduction) and is rounded to the next highest whole number

Absolute risk reduction - CER - EER

Experimental event rate (EER) = (Number who had particular outcome with the intervention) / (Total number who had the intervention)

Control event rate (CER) = (Number who had particular outcome with the control/ (Total number who had the control)

Key trials: lipid management

The following table summarises some of the key trials which have altered the approach to lipid management:

The 4S trial	The 1994 Scandinavian Simvastatin Survival Study was a double-blinded randomised controlled trial looking at the secondary prevention of cardiovascular disease. Patients who had ischaemic heart disease and a cholesterol between 5.5 and 8.0 mmol/l were given either simvastatin or a placebo. Main results • total mortality decreased by 30% with death related to ischaemic heart disease decreased by 42% • no increase in non-cardiovascular death
WOSCOPS	The 1995 West of Scotland Coronary Prevention Study was a randomised controlled trial of men aged 45-64 years with no history of ischaemic heart disease and with a raised cholesterol (> 6.5 mmol/l). Participants were given either pravastatin or a placebo. Main results • total mortality decreased by 22% with death related to ischaemic heart disease decreased by 31%
Heart Protection Study	A large randomised controlled trial of just over 20,000 patients funded by the Medical Research Council. Initial results were published in 2002. Patients were included if they were between 40 - 80 years and were considered to have a substantial 5-year risk of death from ischaemic heart disease due to a history of vascular disease or risk factors such as diabetes or hypertension. Patients were randomly allocated either simvastatin 40mg, antioxidants (600 mg vitamin E, 250 mg vitamin C and 20 mg beta-carotene daily), placebo or a combination. Main results • number needed to treat (NNT) to prevent all cause death = 57, NNT to prevent death related to ischaemic heart disease = 85 • NNT to prevent a vascular event = 19, NNT to prevent a major coronary event = 33, NNT to prevent a stroke = 73 • vascular events were reduced by around 25% • antioxidants did not affect clinical outcome

Graphical representations of statistical data

The table below gives a brief summary of the main types of graphs used to represent statistical data.

Box-and-whisker plot	Graphical representation of the sample minimum, lower quartile, median, upper quartile and sample maximum
Funnel plot	Used to demonstrate the existence of publication bias in meta-analyses
Histogram	A graphical display of continuous data where the values have been categorised into a number of categories
Forest plot	Forest plots are usually found in meta-analyses and provide a graphical representation of the strength of evidence of the constituent trials
Scatter plot	Graphical representation using Cartesian coordinates to display values for two variables for a set of data
Kaplan-Meier survival plot	A plot of the Kaplan-Meier estimate of the survival function showing decreasing survival with time