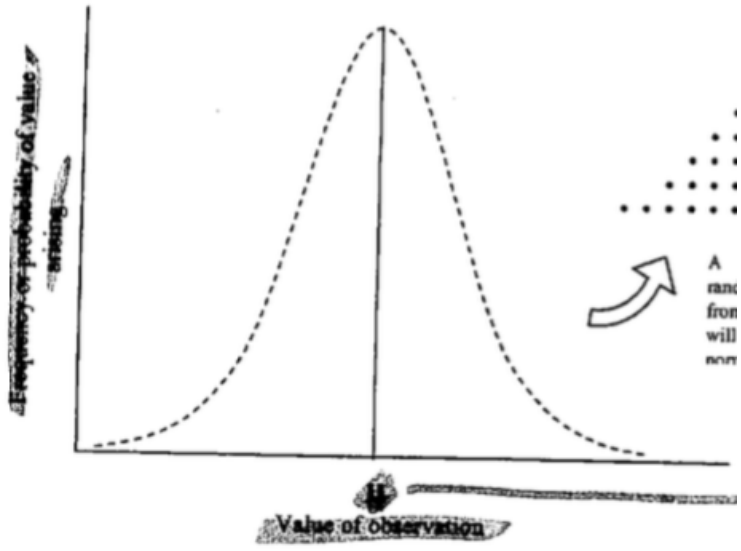
Contents

Basics

Types of Data

- parametric = continuous, numerical data from normally distributed population
- non-parametric = other data which is not normally distrubuted …. e
  - ordinal data eg number give to subjective observations eg ASA, APGAR
  - nominal data eg variable described in terms of quality not quantity eg frequency of waveforms
  - interval data = like ordinal data but intervals between values is equal but doesnt start at zero eg temp in celcius (20deg is not twice as hot as 10deg)
  - ratio = interval data with a natural zero point eg temp in kelvin or absolute pressure
  - binary data = 2 alternatives eg dead or alive
  - discrete = data isolated & separated by gaps eg number of episodes of vomiting
  - continuous = data part of continuous range of values eg height
- reporting data
  - parametric ⟹ mean & standard deviation
  - interval & ratio ⟹ debate whether mean appropriate
  - ordinal ⟹ median, range
  - nominal ⟹ mode

Normal Distribution



Features

- observation within population has a norm
- random independent factors cause variation on that norm:
  - equal spread of values about and below norm
- most values cluster around norm
- ↓ing values seen further from norm
- extreme values do exist though
- mean (average value) = median (central value) = mode (most common value)
- plotted curve =
  - bell shaped as above
  - tails never reach x axis
- number of samples:
  - n>100 usually = near normal
  - smaller numbers ≈ ↓likely normal distribution
- width of curve = measure of variability

- SD is a measure of this

Is it Normal

- plot values & eyeball
- calculate SD:
  - normal distribution must
    - = 68.3% of results within 1 SD of mean
    - = 95% of results in 1.96 SD of mean

Indices of Central Tendency

- Arithmetic mean = sum of observations divided by number of observations
  - Used for ratio or interval data
- Median = middle value of a series of observations
  - Ordinal data
- Mode = Value which occurs most frequently
  - ordinal data
- Geometric Mean = used if data transformed to logs of significance testing

Measures of Variability

- describe average dispersion of data around a mean
- terms:
  - range:
    - = smallest & largest value in a sample
    - heavily influenced by outliers
    - commonly used in reporting non-parametric data
  - percentiles:
    - what percentage of scores is less than your one
    - eg 65th percentile = 65% less than you
    - 50th percentile = median
    - interquartile range = middle 50% of observations
  - standard deviation:
    - = measure of average spread of individual values around sample or population mean
    - to calculate:
      - square differences between each value & sample mean
      - sum the squares
      - divide this by n-1 ⟹ gives variance
      - then square root variance

$$SD = \sqrt{\left(\frac{\sum (x_i - \bar{x})^2}{n-1}\right)}$$

  - degrees of freedom:
    - (n-1) = degrees of freedom
    - is number of independent observations which are possible in a sample
    - it is one less than sample number as the last one can be deduced

Standard Deviation

- Benefits of standard deviation:
  - SD with mean gives indication as to whether mean represents a real trend in sample
  - if large randomly selected sample then SD likely close to mean
  - SD can be used to calculate standard error of the mean (SE)
  - data points within normal distribution can be described as however many SDs from mean
    - tables then tell you proportion of values more extreme than that
    - = z transformation

Standard Error of the Mean

- = estimate of spread of sample means around a population mean
- is estimated from data in single sample
- useful:
  - used in parametric tests to quantify difference between a sample mean & it's

$$SE = \frac{SD}{\sqrt{n}}$$

proposed population mean
- used to calculate confidence intervals

Parametric Tests

- based on parameters of the normal distribution
- determine likelihood of a difference occurring by chance variation rather than real effect
- assumptions:
  - data is continuous & numerical
    ↳ can treat large numbers of discrete data as parametric
  - samples have same variance
  - taken randomly from normally distributed population
- null hypothesis:
  - = any apparent difference or effect is a random variation of no difference or effect
  - hypothesis test is carried out to determine likelihood of this statement being true or false
- p value:
  - proportion of the "standard normal distribution curve" which is more extreme than the z value
    ↳ = curve created which always has mean of zero & SD of 1
  - ∴ probability that difference has occurred by random variation alone
  - p value gives likelihood of null hypothesis being correct
- alpha value:
  - significance level set at study design stage
  - = limit at which p value too large for difference to be regarded as statistically significant
  - in med research = 0.05
- comparing p to alpha:
  - p 0.045 = 4.5% chance difference found occurred by chance alone
  - problems:
    - 0.05 alpha fairly arbitrary
    - statistical significance does not = clinically significant difference
    - ↑ed alpha value = ↑chance of false positive error (aka alpha or type 1 error)
- type 1 error (alpha error):
  - frequency where we in error conclude there is a difference when there isnt one ie false positive frequency
- type 2 error (beta error):
  - frequency where are unable to detect a difference when there is one ie false negative frequency

- one tailed vs two tailed hypothesis:
  - research aim phrased:
    - A is different from B
    - is A larger than B
  - better to have totally open mind about potential direction = 2 tailed hypothesis
  - if absolutely know beforehand which direction variance will occur can use 1 tailed hypothesis
    ↳ is easier to achieve statistical significance with 1 tail

Student's T test

- parametric test for means of samples which are from a normally distributed population but which are 2 small for z test
- diff types:
  - 1 sample t test = likelihood of a sample mean being different from a specified number
  - 2 sample or unpaired t test = likelihood of means of 2 independent samples being different
  - paired t test = likelihood of 2 sample means being different where the samples are the same individuals before & after an intervention

Confidence Intervals

- = range around sample mean within which you predict the mean of the sample population lies
  ↳ ie range within which you predict true value lies

- 95% of sample means lie between 1.96 SEM above & below population mean
- gives an indication of precision of sample mean as an estimate of population mean
    ↳ wider interval means greater imprecision ∴ greater potential difference between calculated sample mean & 'true' mean
- causes of wide CI's:
    - small samples
    - large variance in sample
- p compared to CIs:
    - p value = probability of specific hypothesis being right or wrong (binary)
    - CI =
        - ↑ed scope for reader judgement on significance (graded)
        - overlapping CI's cannot be regarded as different
- odds ratios:
    - frequently presented with CIs
    - OR = 1 ≈ no risk associated with exposure
    - if OR with CI including 1 than OR cannot be significant
- pooled OR of meta-analysis:
    - presented as a diamond at bottom of forest plot
    - width of diamond = CI
    - if diamond crosses vertical line (OR =1) then pooled OR is not significant

Analysis of Variance (ANOVA)

- determines whether difference among three of more samples by comparing
    - variability between groups (should be large)
    - variability within groups (should be small)
- type:
    - one way ANOVA = compare 1 observation in 3 or more groups
    - multiple anova = compare >1 observation in ≥3 groups
    - repeated measures ANOVA = comparing 1 variable in same group at different times

Non-Parametric Tests

- appropriate when:
    - distribution of data is severely non-normal
    - ordinal or discrete quantative data
    - small samples
- characteristics of non=parametric tests:
    - based on ranking
    - results are reported with median & range rather than mean & SD
    - less powerful than parametric tests - type II error more common (false negatives)
- assumptions:
    - samples randomly selected
    - observations independant

Wilcoxon Rank Sum Test

- = non-parametric equivalent to unpaired T test
- principle:
    - 2 samples combined, ordered & ranked from low to high
    - samples seperated & ranks summed
    - look to see if difference between sums of 2 groups

Mann-Whitney U Test

- = non-parametric equivalent to unpaired T test

- principle:
  - rank all smples from smallest to largest & sum rankings in each sample
  - U statistic calculated to assess likelihood of a difference between rank sums
  - U statistic is located in U probability tables

Linear Regression & Correlation

- used to compare relationship between 2 variables where relationship is continuous eg bp & haemorrhage
- linear regression = drawing of a line which best describes association between variables
- correlation = closeness of association between the 2 variables
- assumption:
  - relationship is linear
  - observations are independant
  - observations normally distributed

Pearson Correlation Coefficient

- correlation sis assessment of how likely proposed relationship is
- based on quantifying residual scatter around the regression line
- r value:
  - 1 or -1 = perfect correlation
  - 0.7-1 = strong correlation
  - 0 = no correlation

Risk

Chi square vs Risk Analysis

- Chi square assesses likelihood to be real numerical difference in frequency if an event between groups
- Risk analysis:
  - = gives indication of the strength of association between groups
  - several ways to score risk:
    - relative risk
    - odds ratio
    - number needed to treat

Cohort Studies

- = study of patients where some exposed to risk, others not
- followed over time to determine which develop disease
- almost always prospective (but is poss to do retrospective)
- most commonly use relative risk (OR also possible)

	Cancer	No Cancer
Smoking	a	b
Non-smoking	c	d

Relative Risk (Risk Ratio)

- = ratio of incidence of disease among exposed : incidence among non exposed
- aka incidence risk

$$RR = \frac{\text{incidence among exposed}}{\text{incidence among non - exposed}} = \frac{\left(a/a+b\right)}{c/c+d}$$

- key points:
  - = a true risk ie RR of 3 means x3 the risk
  - RR is reported with CI & if CI includes 1 then not significant

Case Control Study

- = study where cases identified retrospectively as having disease
- then compare those patients with controls without disease
- useful in very rare conditions
- number of cases & controls which had exposure to variable of interest is compared
- analysis with Odds Ratio

	DVT	Controls
OC pill	a	b
No OC pill	c	d

Confounding Variables

- = form of bias seen when demographics of groups studied are different & demographics influence outcome
- eg ages or co-morbidities not matched between groups
- in order to prevent:
  - design stage:
    - large sample
    - randomisation
    - stratum matching - several studies with diff age groups
    - matched design
  - analysis stage:
    - subdivide into diff age groups & analyse separately
    - logistic regression
    - multivariate analysis

Odds Ratio

- odds of disease = number of cases who have disease divided by number who dont have it
- odds ratio = odds of disease in exposed divided by odds of the disease in non exposed

$$OR = \frac{\text{odds of disease in exposed}}{\text{odds of disease in non - exposed}} = \frac{\left(a/b\right)}{c/d}$$

- key points:
  - OR does not give an exact value of risk
  - actually tend to overestimate risk except where outcome is rare
  - reported with a CI
  - only test appropriate for retrospective case-control studies

Mnemonic to remember OR:RR Cohort:Case Control

Backward	coHort
OR	Ahead (ie prospective)
coNtrol	RR
Exposure	Disease

Number Needed to Treat

- = number of patients who need to be treated in order to avoid one adverse event
- NNT is the reciprocal of the absolute risk reduction
- NNT advantages over RR:
  - gives relevance in terms of magnitude of clinical effect - impt in rare problems:
    - incidence of adverse event is rare = 0.06%:

        - 33% reduction in risk (RR = 0.33) ⟹ absolute risk reduction of only 0.02%
        - ∴ NNT to prevent one adverse event would be 5000
    - incidence of adverse event is common = 6%:
        - RR 0.33 ⟹ absolute risk reduction of 2%
        - ∴ NNT of only 50
- calculation of NNT: **NNT = 1/ARR**
- **ARR =** (drugs ↓s risk of bad outcome from 50% to 30%)
    - control event rate - experimental event rate
        - 0.5 - 0.3 = 0.2
        - 0.2 = 20%
- **NNT**
    - ∴ absolute risk reduction = 0.2
    - NNT = 1/0.2 = 5

Predictive Ability of Tests

- diagnostic test:
    - any kind of test performed to aid in the diagnosis or detection of disease

Sensitivity

- = true positives / true positives & false negatives
- = ability of test to detect disease
- = true positives correctly identified by test
- ∴ high sensitivity = ideal

Specificity

- = true negatives / true negatives & false positives
- = true negatives correctly identified by test
- ∴ high specificity = test with few false positives

↳ esp impt in screening tests

↳ can be estimated from case-control studies ie dont need to be able to estimate prevalence pre-test

- ie more useful in disease which are less prevalent

PPV

- = true positives / true positives & false positives
- proportion within population who test positive that actually have disease

NPV

- = true negatives / true negatives and false negatives
- = those who test negative that don't have disease 'true negative'

↳ Predictive values depend on prevalence of disease and may vary from population to population:

- need to know estimates of prevalence from cross sectional studies
- ie much better high when prevalence of disease is more common
- if disease is very uncommon, would need to have a very very high NPV to say someone doesnt have a disease

- Likelihood ratio for positive test result (LR+) = sensitivity / 1 – specificity
- Likelihood ratio for negative test result (LR-) = 1-sensitivity/specificity
- Posterior odds = prior odds multiplies by likelihood ratio

		Condition (as determined by "Gold standard")		
		Condition Positive	Condition Negative	
Test Outcome	Test Outcome Positive	**True Positive**	**False Positive** (Type I error)	**Positive predictive value** = Σ True Positive / Σ Test Outcome Positive
	Test Outcome Negative	**False Negative** (Type II error)	**True Negative**	Negative predictive value = Σ True Negative / Σ Test Outcome Negative
		Sensitivity = Σ True Positive / Σ Condition Positive	Specificity = Σ True Negative / Σ Condition Negative	

Power & Calculation of Sample Size

- unethical & waste of time & money to embark on study to see if drug is effective if there is a significant chance of false negative result
- most commonest cause of this is too small sample size
- power of study
  - = chance of it successfully demonstrating the true result
  - 1- the false negative rate

Requirements to Calculate

- desired effect size - ie effect hoping to demonstrate (plus also what might be regarded as no effect ie null hypothesis value)
- power size
  - ie how certain you want to be in picking up true effect
  - conventionally we use a power of 80-90%
  - high power sizes ⟹ larger sample size required
- alpha value:
  - by tradition 0.05
  - lower values ⟹ ↑sample size
- prediction of variance within samples:
  - commonly taken from pilot studies or literature studies
  - larger variance ⟹ larger required sample

- final calculations used depends on type of trial
- complex equations

Meta Analysis

- = mathematical process of combining numeric data from studies using similar treatments
- done in a systematic manner
- systematic review = the process of collecting studies, meta-analysis, commenting & conclusions

Aims & Advantages
- pooled estimate of effect
- allows for objective appraisal of evidence
- may ↓probability of false negative results
- heterogeneity between study results may be explained

Problems
- heterogeneity of study demographics, methods, results & quality
- selection bias of studies & data
- use of summary data rather than individual data ⟹ magnification of assumptions/errors
- lack of inclusion & exclusion criteria detail
- publication bias ⟹ many negative studies not published

Practicalities
- different studies are weighted:
  - should be done in transparent manner
  - largest trials most heavily weighed
- OR used & combined using random effects model
- Forrest plot used to graphically display
  - OR & confidence intervals
  - pooled OR
- positive MA should always be confirmed with a large RCT

Funnel Plots
- random variation ⟹ spread of study results around the true result
- larger studies will be closer to true result
- ∴ plot of result against size of all studies in MA should create a symmetrical funnel shape
- if not symmetrical ≈ publication bias
- need number of studies for this to be accurate
- eg of funnel plot ≈ bias……..

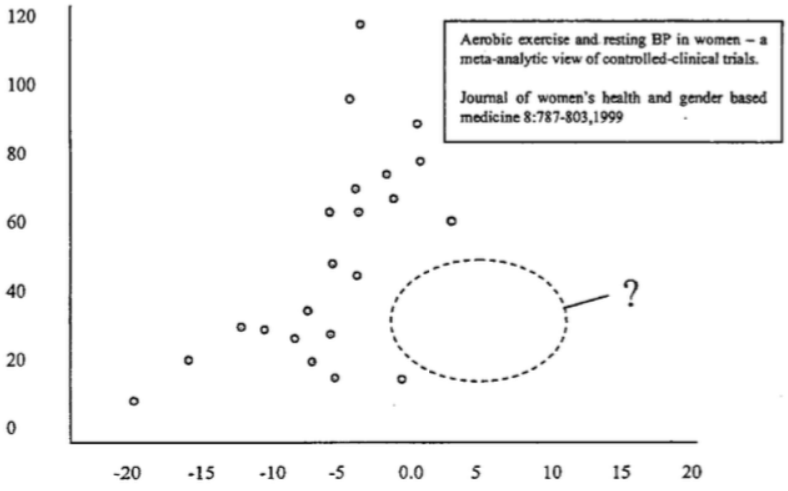


Figure 18: The Funnel Plot with a hole in right side suggesting missing studies

Evidence Based Medicine
- = application of current best evidence in the management of individual patients
- phases to consider:
  - 1 = ask solvable question
  - 2 = research type, quality
  - 3 = is evidence valid
  - 4 = evidence applicable to your patient

- 5 = self assessment

Quality of Evidence

many diff systems but 1 example….
- Level 1 = systematic review of all relevant RCTs
- level 2 = at least 1 well defined RCT
- level 3 = other well designed trials
- level 4 = descriptive studies, reports of expert committees or opinions of experts

Categories of Recommendation

- based on balance of risk versus benefit to patient:
  - level A = good evidence suggest benefit substantially outweigh risk
  - level B = fair evidence suggest benefit outweigh risks
  - level C = fair evidence that benefits (eg to treatment). but balance of benefit & risk is too close to make general recommendation.
  - level D = fair evidence that risks outweight benefit.

↳ level A & B - Treatment should be offered to pt
level C - treatment offer = judgement call
level D - treatment should prob not be offered unless extenuating circumstances

Errors in Research

Random Error

- = error from lack of precision in conducting study
- ↓ed by meticulous technique & by studying large numbers

Bias

- = introduction of systematic error
- not ↓ed by ↑ing sample size

Examples

Problem	Explanation	Prevention
Selection bias	- 1 gp has different risk than the other	- randomisation - Cross over
Detection bias	- observations in 1 gp not sought as diligently as in other	- Blinding
Recall bias	- Pts allocation group influences way they report symptoms eg placebo vs treatment	- patient blinding
Response bias	- Pts enrolling in a trial may not represent population as a whole	- random selection
Publication bias	- Negative studies less likely to be published	- all studies should be submitted regardless of result - MA's should demonstrate funnel plot analysis
Regression to mean	- Random effects ⟹ rare, extreme variation on a measurement - if measurement repeated then measurement likely less extreme - ∴if treatment given after 1st measurement then repeated 2nd measurement may falsely suggest a treatment effect	- Control group

Problem	Explanation	Prevention
Hawthorne Effect	- Actual process of studying & following pts influences outcome	- Control group - mask intention of study from pt
Confounding	- when demographics of groups are different	- Match groups - analysis in stratifications
Sample size too small	- ↑risk type 2 error	- adequate sample size

Analysis Errors

- parametric test used rather than non-parametric:
  - common if:
    - popn is not normally distrubuted
    - sample size is too small to be sure if it is of normal distrubution
    - ordinal data treated as interval data
- opposite (non-parametric used instead of parametric):
  - non parametric tests are less powerful
  - ↑ed risk of type II error (false negative)
- paired data treated as unpaired ⟹ ↑ed chance type 2 error
- one tailed test instead of 2 tailed ⟹ ↑ed chance of type 1 error

Presentation/Publication Error

- failure to report data points or SD or SE
- reporting mean with SE:
  - should report with SD
  - SE is always a smaller number ∴ false impression of trend in sample
- assumption that a p value less than alpha value = clinical significance
  - nope - it only suggests statistical significance
- failure to publish study design & statistical analysis
- publication bias

How to Plan a Study

- define aim
- research topic
- write protocol:
  - aim
  - background
  - study design:
    - prospective vs retro
    - cohort vs case control
    - sequential trial design
  - Reducing bias:
    - blinidng
    - crossover
    - randomisation
    - controls
  - inclusion/exclusion criteria
  - power analysis - calculate sample size
  - treatments: dose, side effects, placebo
  - outcome measures

    - statistical methods & null hypothesis
    - safety monitoring
    - patient info & consent forms
- ethics approval
- pilot study
- modify protocol based on basis pilot
- re-ethics approval
- execute

Clinical Drug trials

- pre-clinical = non human subjects to understand efficacy, toxicology, pharmacokinetics
- phase 1 =
    - dose ranging
    - subtherapeutic dosing
    - humans (20-100) pharmacokinetics & toxicology of drug
- phase 2 =
    - assess efficacy & safety
    - therapeutic dosing
    - humans (100-300)
    - not looking for therapeutic effect
- phase 3 =
    - assessing effectiveness
    - humans 1000-2000
    - testing for therapeutic range, assumed to be effective
- phase 4 =
    - post marketing surveillance
    - watch long term effects

High Yield Definitions

Normal Distribution
- random independant factors have caused a spread of observations around a norm
- most values are around norm
- extreme variations are rare
- random effects work above & below norm ∴ mean = mode = median
- bell shaped plot
- large sample randomly from normal population also has normal distribution
- big sample sizes = mean & SD likely to be close to population mean
- small sample sizes ⟹ ↑chance mean & SD further from population mean
- if multiple samples taken from same populatio then plot of means will be norm distributed

Standard Deviation
- = spread of individual values around population mean
- Use:
  - gives indication of reliability of mean as a summary of trend in sample
  - SD of large sample is similar to that of its population
  - Used to calculate SEM
  - used for z transformation of individual observations

Standard Error of Mean
- measure of spread of sample means around the population mean

Standard Error
- used in z transformation of sample means in parametric testing
- used to calculate confidence intervals

Parametric Testing
- tests based on normal distribution
- informs whether difference due to chance or real

Z value
- expresison of an individual observation or sample mean in multitudes of SD or SEMs from population mean

Null Hypothesis
- a hypothesis to be tested which states no real difference between two values ie any difference has occurred by chance

Alpha
- = significance level chosen at study design stage
- = limit at which p becomes statistically significant

P value
- = probability of there being no difference when you say there is
- = probability that difference has occurred by random variation alone
- = gives likelihood of null hypothesis being correct
- calculated from trial results

Alpha (type 1 error)
- chance of there being no difference when you say there **is** one
- false positive rate

Beta (type II error)
- chance of being a difference when you say there **isnt** one
- false negative rate

Confidence interval
- range above and below the sample mean within which you predict the sample population mean lies

T Test
- parametric test of means where samples are too small to use the normal test

One tailed
- When there is only one direction that one group can vary from another
- ∴ only have to look for one tail
- = easier to get significant result ∴ if used incorrectly ↑ed chance type I error

Two Tailed
- dont know for certain which way test result will vary
- ∴ look for two tails
- = harder to get significant result ∴ used incorrectly ↑chance of type II error

ANOVA
- method used to compare 3 or more parametric samples
- between group variance must outweigh within gp variance

Non-Parametric Testing
- any test which isnt parametric
- uses ranking & data which isnt continuous

Regression
- drawing line which best describes relationship between 2 continuous variables

Correlation
- how close relationship is between variables

Power of Study
- = probability of a study being able to demonstrate a difference when a difference exists
- 1 - false negative rate

Calculating sample sizes
- comparing numbers in 2 samples:
    - alpha
    - 1-power
    - desired effect size
    - predicted standard deviation of samples
- comparing proportions:
    - alpha
    - 1-power
    - proportion looking for
    - null hypothesis proportion

Chi Square
- compares frequency of binary event within 2 or more groups
- uses a contingency table
- compares observed with expected values

Relative Risk
- incidence of an event with exposure compared to without exposure

Odds Ratio
- odds of getting an event with exposure compared to without exposure

NNT
- number of patients needed to be treated to avoid an adverse event
- 1/ARR

Sensitivity
- proportion of disease correctly identified
- true positive rate identified by test

Specificity
- Proportion of no-disease correctly identified
- ie true negative rate identified by test

PPV
- proportion of a tests positive results which are true positives
- must have prevalence rates

NPV
- proportion of tests negative results which are true negatives

Meta Analysis
- mathematical process of combining data from studies using similar treatments in a systematic manner