Statistics

Types of Data (CONIR)

- Categorical: groups — Mode
  - Binary (eg. male/female)
  - Nominal (eg. A, B, AB, O blood types)
- Ordinal: no fixed interval — Median, mode, range
  - Ordinal data (eg. APGAR scores, ASA scores)
- Numerical: Mean, median, range, (mode), SD/variance
  - Interval: fixed interval, cannot multiply, no fixed 0, can add (eg. temp in Celsius)
  - Ratio: fixed interval, meaningful distribution, can multiply and add (eg. temp in kelvin)

- Parametric
  - Continuous
  - Numerical
  - Normal distribution
- Non-parametric

Measures of Central Tendency

- Categorical:
  - Mode: the value that occurs most frequently
- Ordinal:
  - Median: middle of a series of observations; summary of choice in non-parametric data
  - Mode
- Numerical:
  - Mean: (sum of observations)/(number of observations)
  - Median
  - Mode

Measures of Variability

- Categorical:
  - None
- Ordinal:
  - Range: smallest and largest values in a sample
  - Interquartile range
- Numerical:
  - Standard deviation, variance
  - Range
  - Interquartile range

Standard Deviation

- Measure of the average spread of individual samples around the sample or population mean
- +/- 1 SD = 68.3%
- +/-1.96 SD = 95%
- SD = √(variance)
- Variance = $\varepsilon(x-\dot{x})^2/(n-1)$
  - $x-\dot{x}$ is the difference between each value and the sample mean
  - n-1 is the degrees of freedom

Standard Error of the Mean

- An estimate of the spread of sample means around the population mean
- **SE = SD/√n**
- SE is used to calculate **confidence intervals**.
  - Confidence interval is the range around a sample mean with which the mean of the sample's population lies
- This is an indication of the precision of the sample mean as an estimate of the population mean

Confidence Intervals

- The range around a sample mean with which the mean of the sample's population lies

- Indication of precision of the sample mean as an estimate of the population mean
- 95% CI:
    - The range of values in which the true population parameter will fall into, 95% of the time.
    - A large range of values is less significant and relevant than a small range of values
    - Gives clinical significance (not just statistical significance)
- **CI = mean +/- SE x 1.96**
- How this compares to the P-value:
    - P value is the probability of incorrectly rejecting the null hypothesis (false positive)
    - If P = 0.05, there is a 1 in 20 chance of this occurring.
    - It only indicates a statistical significance, not a clinical significance.

Parametric Testing

- Based on the parameters of the **normal distribution**
    - Values cluster around the norm with fewer and fewer values towards the tails
    - Can be completely described by its mean and SD
    - Equal spread
    - Bell shaped curve, symmetrical

Null Hypothesis

- A standpoint adopted in statistical analysis that any apparent difference or effect is simply **a random variation of no difference or effect**.
- Hypothesis test is carried out to determine the likelihood of the null hypothesis being correct.

P-Value

- The probability that the observed difference occurred by chance (random variation) alone
- It gives the likelihood of the null hypothesis being correct
- This is the **acceptable false positive rate**.

α Value/Type I Error

- ***False positives***
- Set by investigators during the design phase.
- It is the limit at which the P value is deemed too large for a difference to be regarded as statistically significant.
- Arbitrarily 0.05.
- This needs to be lower than the β value; the ramifications of rejecting the null hypothesis and saying that there is a difference, and potentially changing treatment/management erroneously is worse.

β Value/Type II Error

- This is the ***false negative rate***
- That is, saying there is not a difference when there is one.
- Usually set at 20%
- Decrease sample size, increase type II error.

Power

- The chance of a test successfully demonstrating the true negative result
    - The ability to detect a difference if there is one
- Power = 1 – false negative rate = 1 – β error
    - Hence there is an 80% probability of detecting a difference if it exists
- Determinants of power:
    - Sample size
    - Magnitude of effect: smaller difference, larger sample size
    - β value: usually 0.2 (if lower, need larger sample size)
    - α value: usually 0.05 (if lower, need larger sample size)
    - Sample variability (if increased, need bigger sample)
    - Number of measured outcomes: sample size might be different for each
    - Sensitivity of the test used
    - Drop out rate

Parametric Tests

	1 – 2 samples	**≥3 samples**
n > 100	Normal test	ANOVA
n < 100	T test	T test with Bonferroni's correction

- Continuous numerical data
- Assumes a normal distribution
- Normal test (z test):
    - n > 100
    - One sample: sample vs population
    - Two samples: sample A vs sample B
- Students' T test:
    - n < 100
    - One sample
    - Two sample unpaired
    - Two sample paired (before and after intervention)
        - Less intrinsic variability
        - More powerful
        - False negative less likely
- ANOVA (Analysis of Variance)
    - n > 100
    - Determines if there is a difference between **3** or more samples by comparing variability
- T-test with Bonferroni's correction:
    - Correction factor allowing a t-test to be used to make comparisons between 3 or more samples

Non-Parametric Tests

- Used when:
  - Non-normal distribution
  - Ordinal or discrete data
  - Small sample
- Compared with parametric tests:
  - Only derive median and range (not mean and standard deviation)
  - Less powerful
  - More type II error
- Wilcoxon Rank Sum
  - Non-parametric equivalent to the unpaired T test
- Mann-Whitney U
  - Non-parametric equivalent to the unpaired T test
- Wilcoxon Paired-Sample Test
  - Non-parametric equivalent to the paired T test
- Kruskal Wallis:
  - Non-parametric equivalent to one-way ANOVA

Categorical Data: Chi-Square

- A test to assess whether there is a real difference between the frequency of a categorical event between 2 or more groups
- Involves a contingency table; compare the observed frequencies with those expected if there is no difference between groups
- Gives a risk analysis
- Relative risk:
    - Ratio of **incidence of disease** in **exposed vs. unexposed**
    - Incidence amongst exposed / Incidence amongst unexposed = a(a+b) / c(c+d)
    - True risk
    - Reported with confidence intervals
    - Seen with cohort studies (ie. Prospective data)

	Disease	No Disease
Exposure	a	b
No Exposure	c	d

- Odds ratio:
    - Odds of the disease in exposed vs unexposed
    - Odds of disease in exposed / Odds of disease in unexposed = a/b / c/d = ad / bc
    - **Not** an exact risk
    - Tends to **overestimate** the risk
    - OR and RR are similar if **rare** outcomes
    - Used in **retrospective** case control studies (no information on numbers of **all** exposed or unexposed

- NNT:
    - Number of patients needed to be treated in order to avoid one event
    - NNT = 1/ARR; where ARR = absolute risk reduction
    - ARR = incidence in exposed – incidence in unexposed
    - *RRR = 1 – relative risk*

	Disease +	Disease –
Test +	a	b
Test –	c	d

Predictive Ability of Tests

- Sensitivity:
  - The ability of a test to correctly detect the disease
  - True positive rate
  - a/(a+c)
  - True positive/Disease
- Specificity:
  - The ability of a test to correctly identify those without a disease
  - True negative rate
  - d/(b+d)
  - True negative/No disease
- Positive predictive value:
  - The likelihood of having the disease if a positive result
  - a/(a+b)
  - True positive/All positive
- Negative predictive value:
  - The likelihood of not having the disease with a negative result
  - d/(c+d)
  - True negative/All negative
- Likelihood ratio:
  - Likelihood of having a positive result when you have the disease compared with the likelihood of having a positive result when you don't have the disease
  - Sensitivity/(1-specificity)
  -
- Sensitivity and specificity:
  - Important in screening tests
  - Not affected by prevalence
- PPV and NPV
  - Take into account the prevalence (prior probability)
  - If a disease is common, a positive result is like to represent a true positive
- Screening tests:
  - High sensitivity (of picking up the disease)
  - High negative predictive value (because prevalence of disease in a large population is low)
- Diagnostic tests:
  - High specificity
  - High positive predictive value

Error

- Random error: error introduced by lack of precision in a study
- Bias: systematic error, not reduced by increasing the sample size
- Types of bias:
    - Selection
        - One group of patients has a different risk to another group
        - Reduced by *randomisation*
    - Detection/observer
        - Observer's assessment is influenced by knowing the treatment allocation
        - Reduced by *observer blinding*
    - Recall
        - Patient's recall of symptoms is influenced by knowing the treatment allocation
        - Reduced by *patient blinding*
    - Response
        - Patient's physical response is affected by knowing the treatment allocation
        - Reduced by *patient blinding*
    - Publication
        - Positive studies are more likely to be published than negative studies
        - Reduced by registration of the trial prior to starting
    - Performance
        - Knowledge of the treatment allocation favours additional interventions or withdrawal
        - Reduced by *allocation concealment* and *observer blinding*
    - Hawthorne effect
        - Being studied makes patients feel better
        - Reduced by *using a control group*.

Critical Appraisal of a Published Article

- Question worth asking?
  - Has it been answered before
- Ethical approval
  - Helsinki declaration:
    - Scientific principles (accepted principles, proper training)
    - Protocols
    - Patients: voluntary, privacy, informed consent, entitled to results
    - Publication: accessibility, conflicts of interest, published regardless of outcome
    - Placebo
    - Adequate risk/benefit assessment
  - Issues in ethical review:
    - Will the trial add knowledge
    - Methodological validity
    - Respect for participants
    - Voluntary informed consent
    - Respect for privacy and confidentiality
    - Potential harm minimised
- Methodology
  - Trial design (RCT, cohort etc…)
  - Participants and selection
  - Sample size
  - Outcome measures:
    - Protocols; were they changed at all
    - Surrogate outcomes?
  - Study design
    - Randomisation
    - Reduction of bias
- Analysis
  - Appropriate?
- Conclusions
  - Justified
  - Spin?
- Applicability and validity
- Conflicts of interest
- Well written

Systematic Review

- Systematic review: addresses a specific clinical question by systematic evaluation of all appropriate trials
- Meta-analysis: the mathematical process of combining all the data from trials from the systematic review
- Protocol:
    - Specific clinical question
    - Inclusion and exclusion criteria
    - Search methods
    - Outcomes
    - Validation of studies
    - Assessment of heterogeneity (diversity amongst study results greater than chance alone):
        - Clinical: different patient groups
        - Methodological: difference in conduction and methods between trials
        - Statistical: consequence of the above two. Significant difference between results
    - Meta-analysis
    - Reliability of results
    - Conclusions
- Requirements:
    - Similar patient groups (↓heterogeneity)
    - Similar interventions
    - Measurement of similar end-points
- Advantages:
    - ↑sample size and power: better chance of detecting a difference
    - Summarises and collates large amount of data
    - More cost effective and time efficient than commencing new trials
    - Reduces delay between research discoveries and starting new treatments
    - Limits bias, improves reliability
- Pitfalls:
    - Publication bias: negative studies are less like to be submitted/accepted/published
    - Duplicate publication: double counting in meta-analysis
    - Heterogeneity
    - Historical/outdated studies
    - Subject to limitations in original study design
    - Results mostly influenced by the results of large trials

Planning a Clinical Trial

- Define the aim:
    - Clinical question
- Perform literature search:
    - Has the question already been answered
- Protocol:
    - Aim
    - Background
    - Design: blinded, prospective/retrospective, crossover, randomisation, control group
    - Inclusion and exclusion criteria
    - Sample size calculation: power (β), α value, size of difference to be detected, patient variability, estimated drop out rate
    - Treatment: dose, placebo
    - Outcomes and data collection
    - Statistical analysis, methods
    - Safety monitoring
    - Patient information and informed consent
- Ethics approval
- Pilot study
- Modify protocol
- Main study