Research, Statistics and Epidemiology

Nov 2017

Antonia Marsden & Matt Gittins

Why medical statistics?

- To use evidence and data to underpin clinical decision-making
- To recognise good evidence and adopt guidelines as appropriate.
- 10% of questions on AKT relate to medical statistics.

Cookie Policy | Feedback Like 13.3M Sunday, Nov 26th

Mail Online

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money | Vid

Latest Headlines | Health | Health Directory | Diets | Discounts

£13million NHS bill for suncream: Millions also wasted on prescriptions for toothpaste, Yakult and Calpol

- NHS gave out 404,500 prescriptions for suncream at a cost of £13m in 2014
- Also handed out 4.7million prescriptions for indigestion pills costing £29m
- Other items routinely prescribed include vitamins, Vaseline and toothpaste
- Critics branded prescriptions 'ludicrous' at time of financial crisis for NHS

By SOPHIE BORLAND, HEALTH CORRESPONDENT FOR THE DAILY MAIL
PUBLISHED: 22:59, 8 April 2015 | **UPDATED:** 17:58, 9 April 2015

Figures based on 2004 Prescription Cost Analysis.

4 out of 43 of the products classed as 'suncream' (not sunscreen) were prescription only medicines for actinic keratoses, with total cost ~£11.6 million.

2 products are available only for people with specific medical conditions, with total cost ~1.2 million.

Leftover cost: ~£200000

Overview of session

- Research design
  - Hierarchy of evidence
  - Study design
- Basic statistical inference
  - Calculating incidence and prevalence
  - Measures of treatment effect
  - Statistical inference - confidence intervals and hypothesis testing
- Screening
  - Sensitivity and specificity
- Questions

Research design

Research design

- Medical studies are typically concerned with investigating and estimating the effect of a treatment/intervention/exposure on some outcome.

Statins

CBT

Smoking

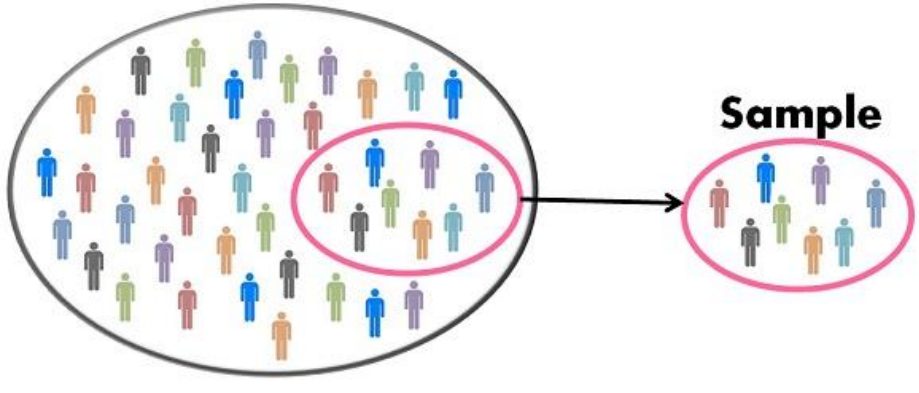
CVD risk

Depression

Onset of rheumatoid arthritis

Research design

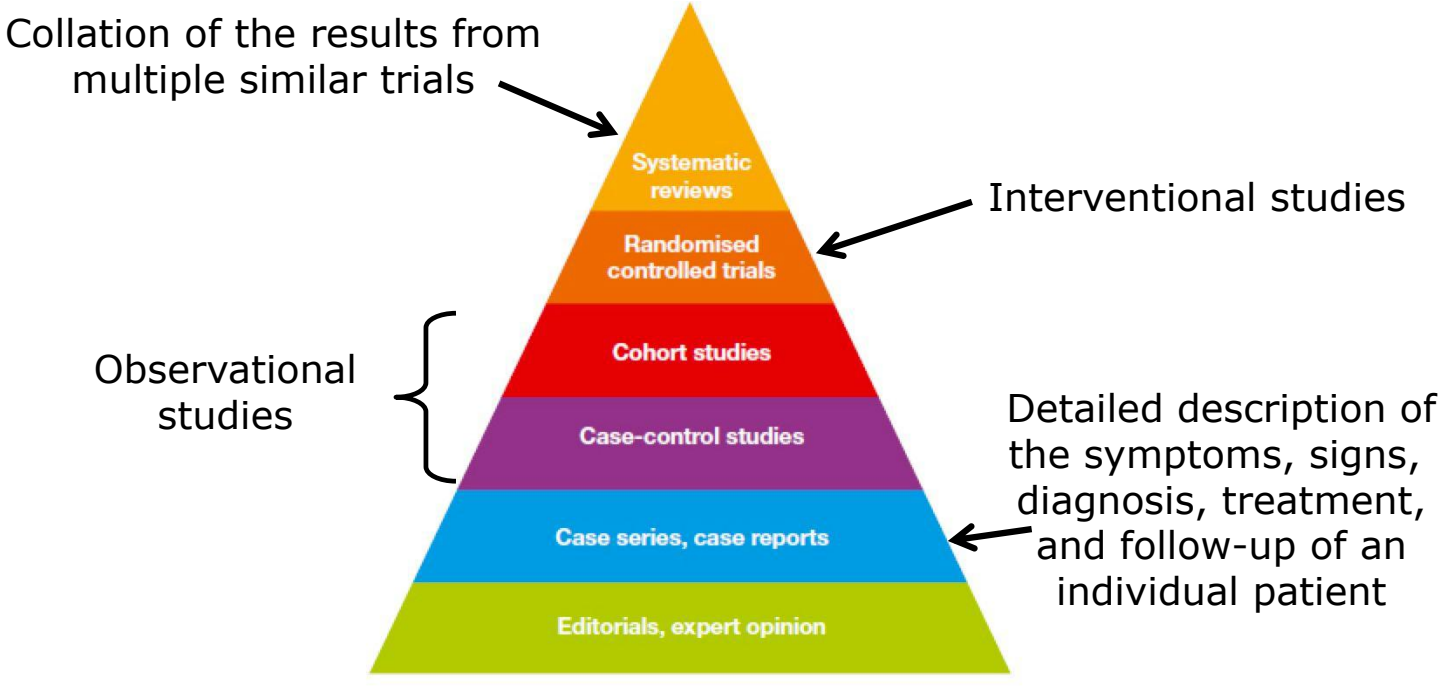
- Evidence needs to come from a **comparison of groups**, e.g. treated vs non-treated, exposed vs non-exposed.
- The non-treated/non-exposed group is referred to as the **control group**.
- Studies are conducted in a **sample** from the population of interest. We use the sample to estimate a population parameter of interest.



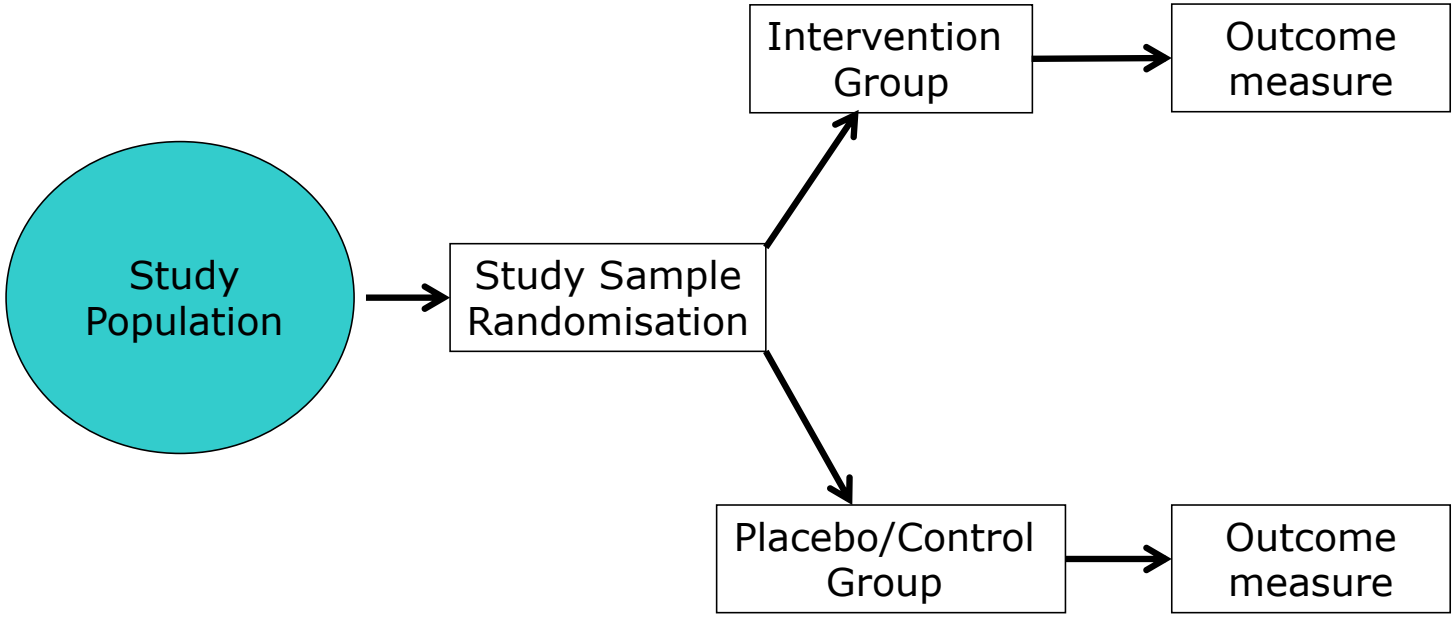
Hierarchy of evidence


Cookie Policy | Feedback
Like 13.3M
Sunday, Nov 26th 2017 9PM 5°C 12AM 7°C 5-Day Forecast
Mail Online
health
Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money | Video | Travel | Fashion Finder
Latest Headlines | Health | Health Directory | Diets | Discounts
Login
Black Friday
offers now live
Terms apply
How using Facebook could raise your risk of cancer
By DAILY MAIL REPORTER
UPDATED: 11:21, 19 February 2009
Share
View comments
Social networking sites such as Facebook could raise your risk of serious health problems by reducing levels of face-to-face
Site
Web
Enter your search
Search
Advertisement
canadian affair
Book your Ontario road-trip today from £999pp
Book now

Hierarchy of evidence



Randomised Controlled Trials (RCTs)



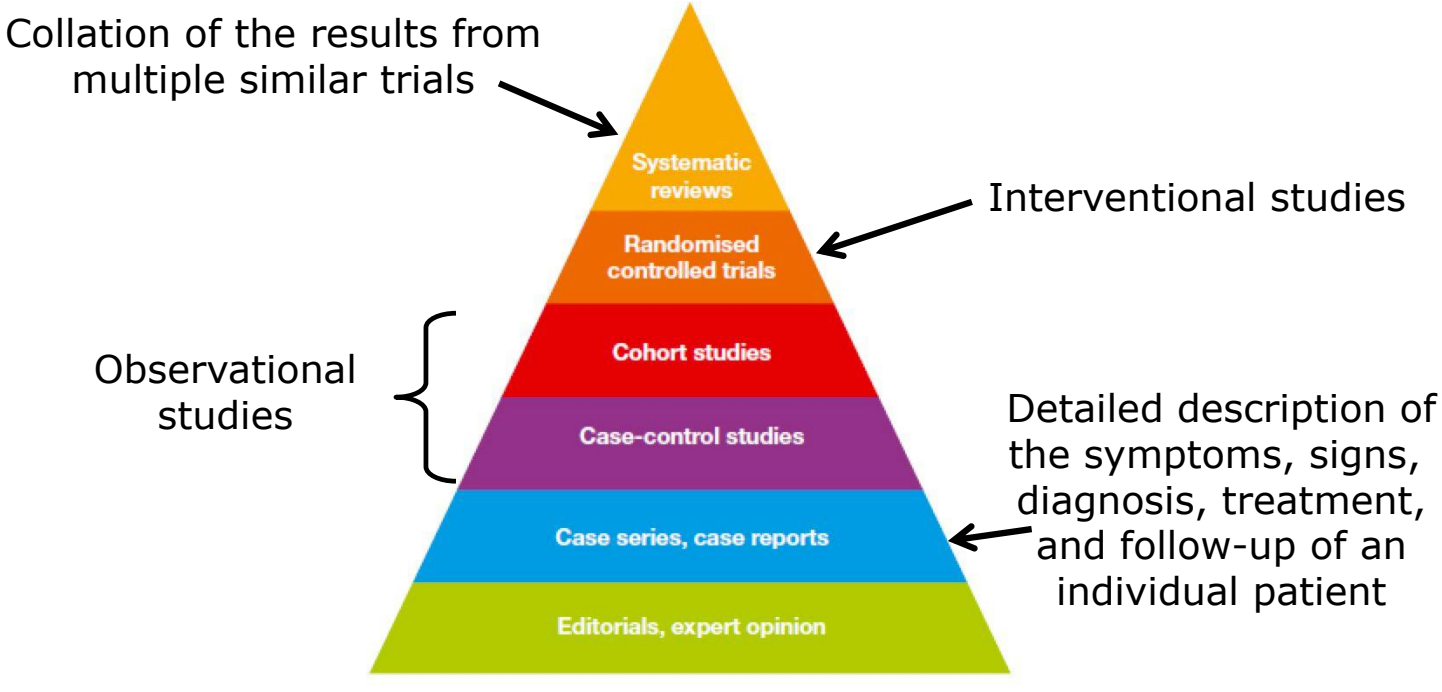
Randomised Controlled Trials (RCTs)

- An RCT is the ‘gold standard’ approach for evaluating the effectiveness of interventions.
- **Randomisation** ensures that the treated and non-treated groups differ only in terms of the treatment receipt – any differences in outcome between the treated and non-treated groups can be attributed causally to the treatment.
- Use of **blinding** and a **placebo** in the control group protects against the fact that people in the treatment group may report better outcomes due to expectations.
- However, trials are expensive, and have ethical, time and sample size restrictions.

RCT: example

- OASIS trial to assess the effect of a digital cogitative behaviour therapy (dCBT) for insomnia.
- 3755 University students were randomly assigned with equal probability to receive dCBT or usual care.
- Measures of insomnia, paranoia and hallucinatory experiences were compared between the two groups at 10 weeks. Compared with usual practice, the dCBT intervention at 10 weeks reduced insomnia (adjusted difference 4·78, 95% CI 4·29 to 5·26) on the Sleep Condition Indicator scale.

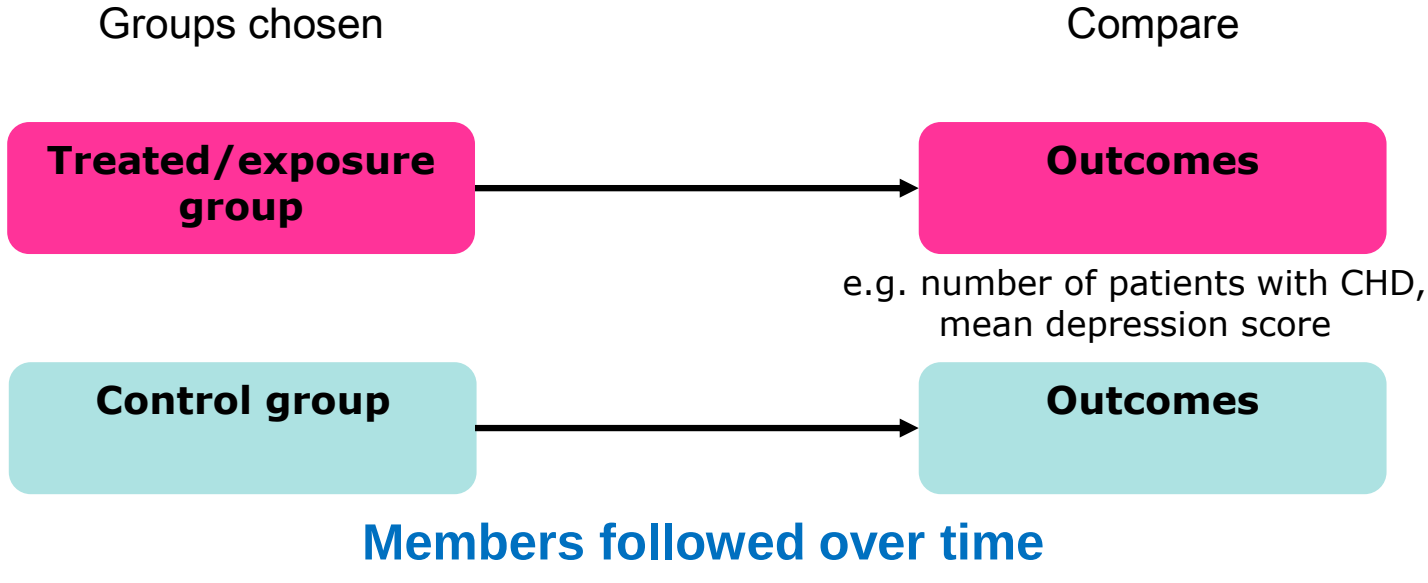
Hierarchy of evidence



Observational studies

- Observational studies are non-experimental studies in which those conducting the study do not influence the interventions received by study participants.
- The two main type of observational study are **cohort studies** and **case-control studies.**

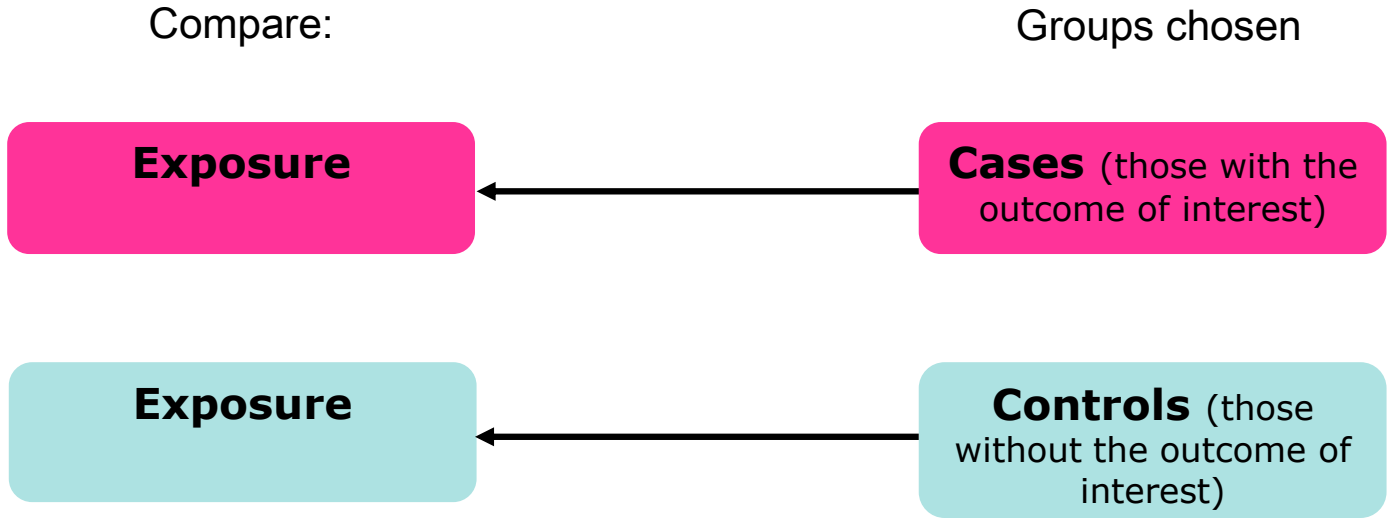
Cohort study design



Cohort study design: example

- The BADBIR is a clinical study to monitor the long term safety of biologic therapies used to treat psoriasis.
- Data is collected on patients who initiate a biologic therapy (via questionnaires) over the years following initiation
- Data on a control group -patients with psoriasis who have not been prescribed a biologic therapy – is also collected.
- A cohort study using data
- from the BADBIR indicated that biologic therapy does not increase the risk of serious infection in patients with psoriasis.

Case-control study design



Look BACK at exposure history in cases and controls

Case-control study design: example

- In the 1980s, the SIDS mortality rate in New Zealand was high (4/1000 live births).
- The New Zealand Cot Death Study was a 3 year case-control study (1987-1990).
- 485 cases (occurrences of SIDS) and 1800 controls were identified.
- This identified three modifiable risk factors for SIDS, namely prone sleeping position, maternal smoking and lack of breastfeeding.
- The number of deaths in the UK fell from 1500 to 600 by the mid 1990s with a concomitant fall in prone sleeping.

Cohort vs case – control studies

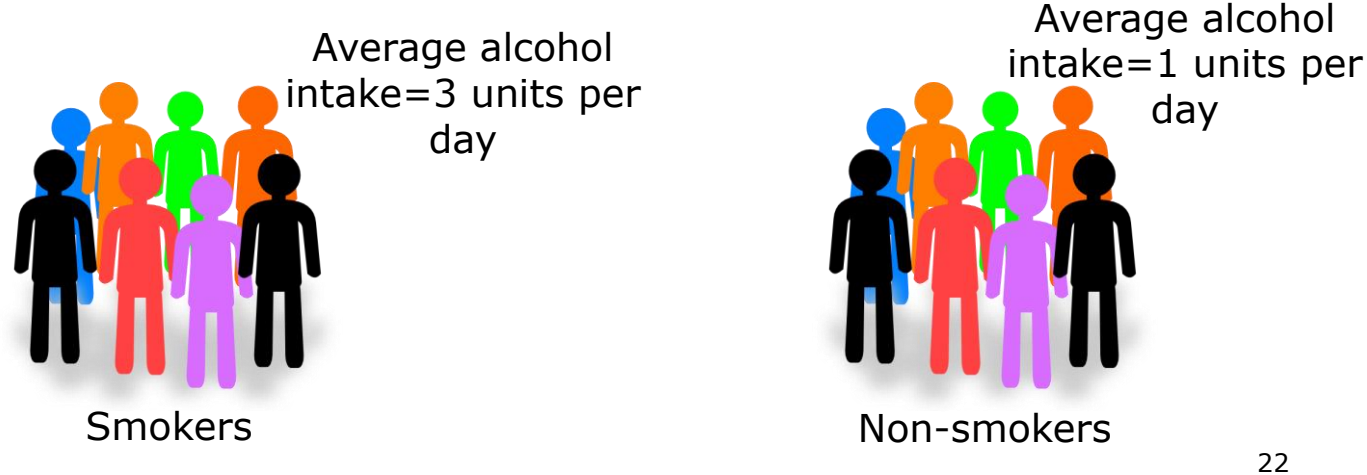
- Case-control are advantageous if the outcome is rare.
- Case-control studies do not give an indication of the prevalence of the outcome (i.e. the percentage of the population who have the outcome).

Prospective vs retrospective studies

- In prospective studies patients of interest are recruited ***before*** any of the subjects have developed any of the outcomes of interest. The subjects are then followed into the future in order to record the development of any of the outcomes of interest.
- In retrospective studies, both exposure status and outcome are ascertained retrospectively.
- You have more control over what data is collected in a prospective study, but they can be lengthy and expensive.
- Cohort studies can be retrospective or prospective.
- Case-control studies are retrospective.

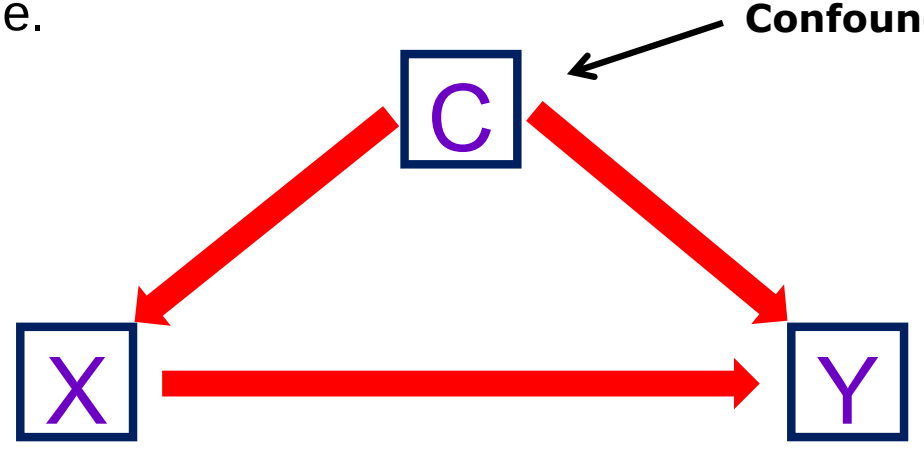
Confounding in observational studies

- The main disadvantage with observational studies is that people who are treated/exposed in practice will likely differ systematically to those who are not treated/exposed and these differences may be also associated with the outcome.



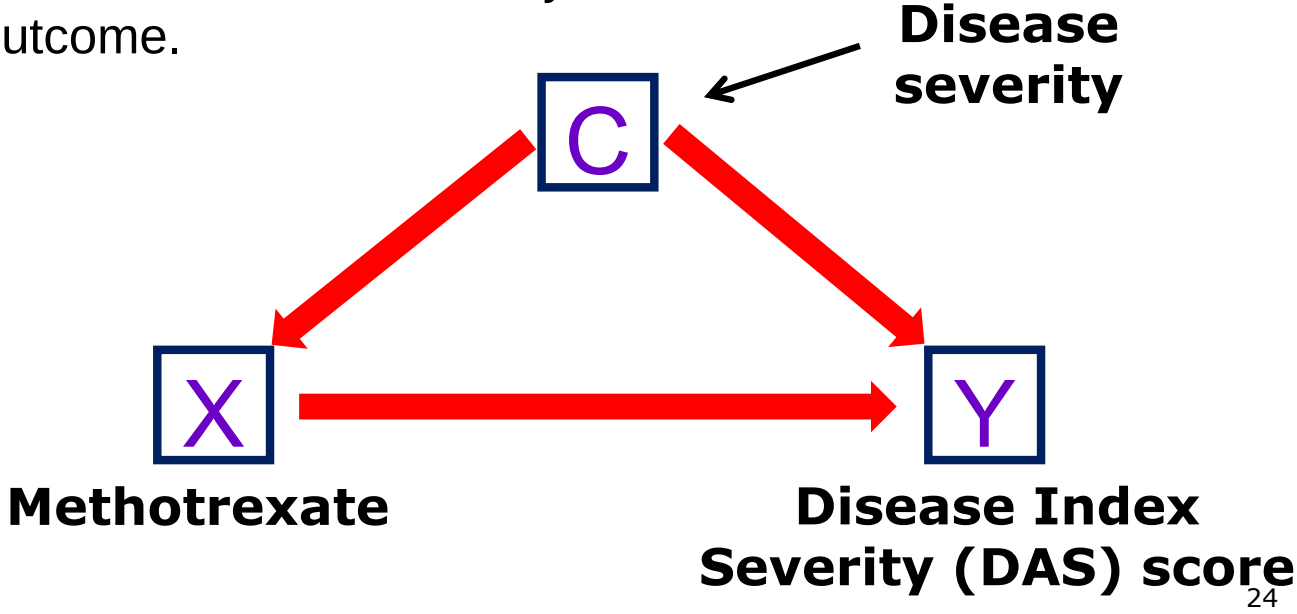
Confounding in observational studies

- The main disadvantage with observational studies is that people who are treated/exposed in practice will likely differ systematically to those who are not treated/exposed and these differences may be also associated with the outcome.



Confounding in observational studies

- The main disadvantage with observational studies is that people who are treated/exposed in practice will likely differ systematically to those who are not treated/exposed and these differences may be also associated with the outcome.



Going to university HALVES your risk of heart disease, 30-year study shows

- Scientists tracked 13,948 white and African-American patients from 1987-2013
- They found graduates had a much lower risk across the US of heart disease than citizens with lower education levels
- Female grads had a 28% lifetime risk compared to 51% of uneducated women
- Male grads had a 42% lifetime risk compared to 59% of uneducated men

By MIA DE GRAAF FOR DAILYMAIL.COM
PUBLISHED: 16:42, 12 June 2017 | **UPDATED:** 16:43, 12 June 2017

16
shares

8
View comments

Confounding in observational studies

- Confounding variables can be accounted for in the statistical analysis of observational studies.
- However, you can't be sure that confounding has been fully accounted for.
- This is why **Randomised Controlled Trials** are performed.

Bias in observational studies

- Bias is 'any trend in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth'.[1]
- Selection bias occurs when there is a systematic difference between those who take part in the study and those who do not.
- Examples include non-response/volunteer bias, invalid choice of control group.
- Information/misclassification bias occurs when there is a systematic reason for us getting the wrong information from individuals included in the study.
- Examples include recall bias, loss to follow-up bias, observational bias.

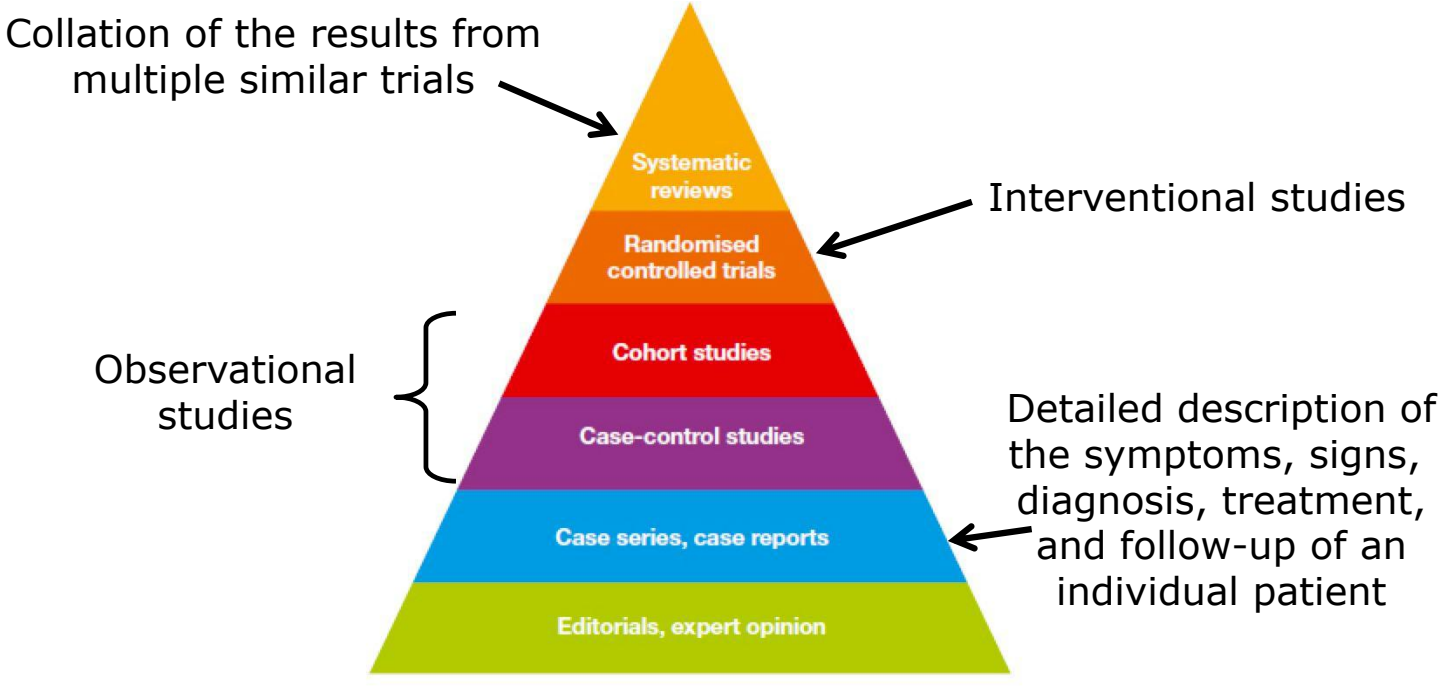
[1] Last, J.M. and International Epidemiological Association (2001) **A dictionary of epidemiology**, 4th ed edn, Oxford ; New York : Oxford University Press.

Exercise

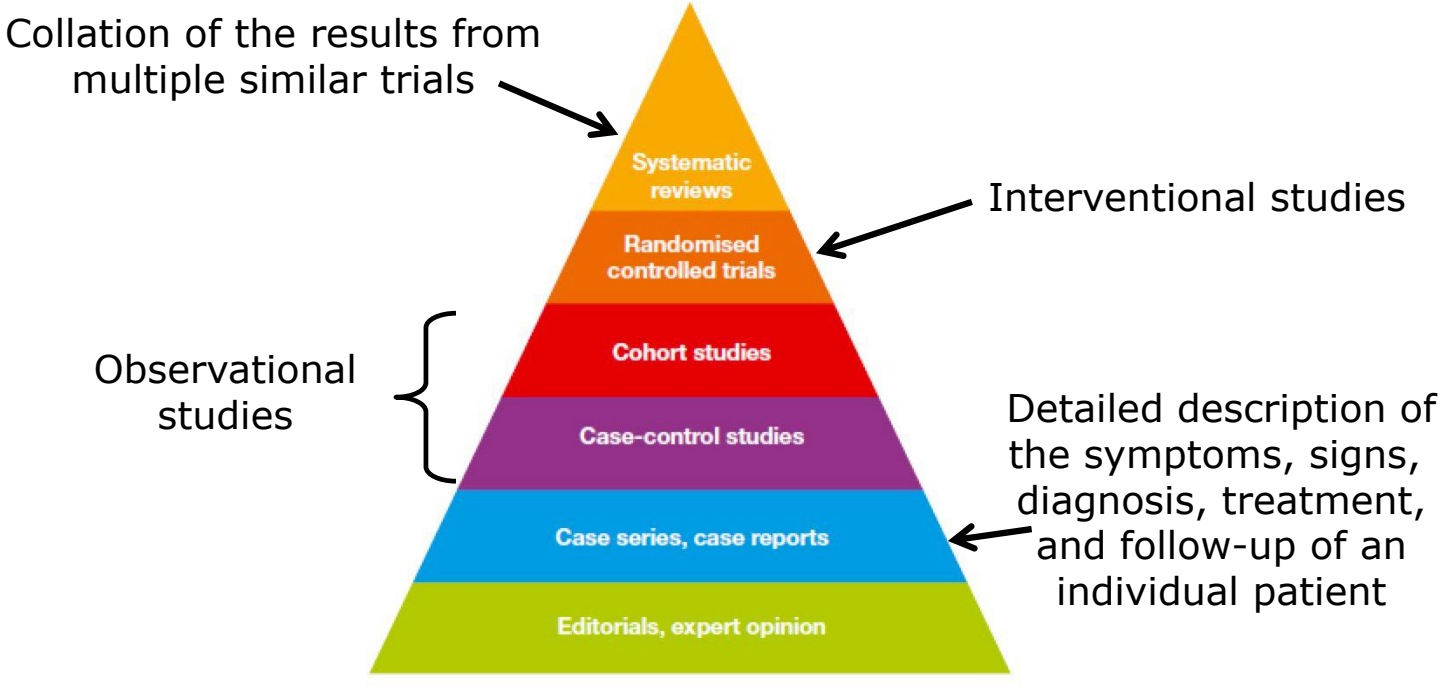
Exercise

- Randomly pick and circle 10 words from the page from Alice in Wonderland.
- Sum together the number of letters in each of your chosen 10 words and divide this number by 10.
- The average number of letters per word on this page is 4.03 – how does this compare to your estimated average?
- Use the random number tables to randomly select 10 words from this page and calculate the average number of letters per word.

Hierarchy of evidence



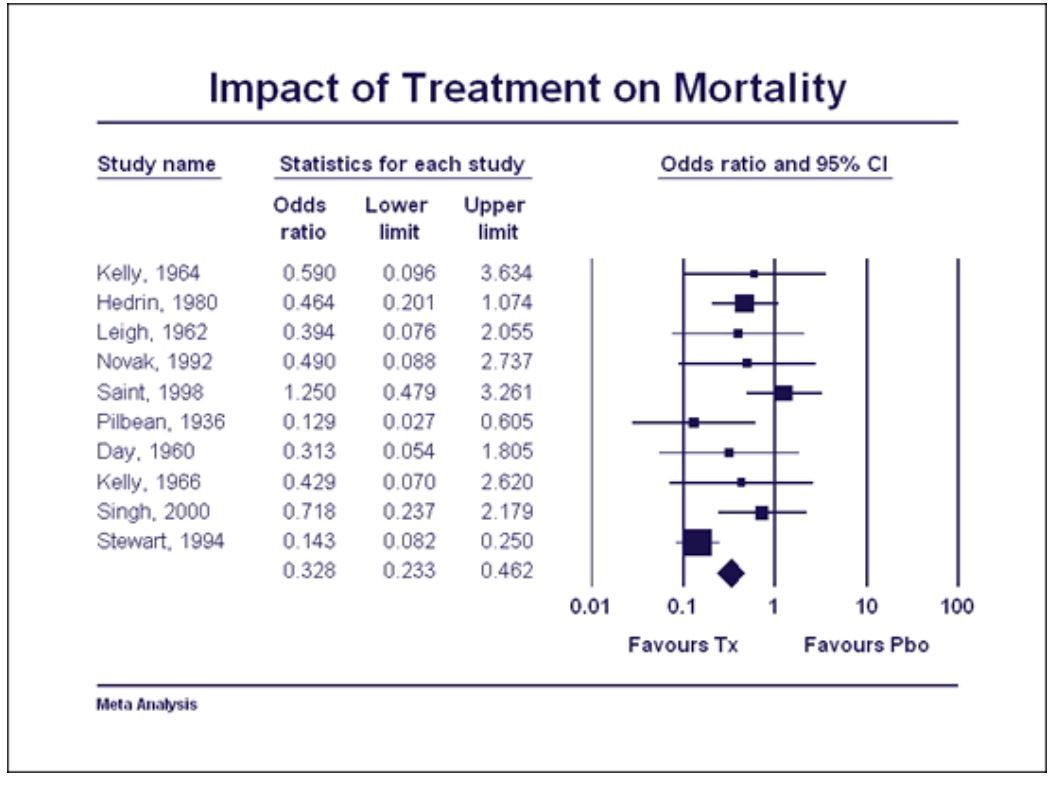
Hierarchy of evidence



Systematic reviews and meta-analyses

- **Meta-analysis** combine the results of multiple RCTs to produce a single treatment effect measure.
- **Systematic reviews** collect and critically analyse multiple research studies or papers looking at a specific clinical question.
- Such studies are useful as the included trials have gone under some sort of critical appraisal by the authors.

Systematic reviews and meta-analyses



Hierarchy of evidence

- **Remember - the quality of evidence relies on the quality of the study!**
- Evidence from a well designed and well analysed observational study will be better than that of a poorly designed trial.
- Trials and observational studies are appropriate in different circumstances – e.g. observational studies are better suited to studying long term and rare outcomes (e.g. rare side-effects to treatments).

Example question

Several studies have reported on the risk of lung cancer and exposure to tobacco smoke. Researchers now wish to reach a summary conclusion about the overall findings.

Which is the SINGLE MOST appropriate study design? Select ONE option only.

A. Case-control study
B. Cohort study
C. Correlation study
D. Descriptive study
E. Meta-analysis

Basic statistical inference

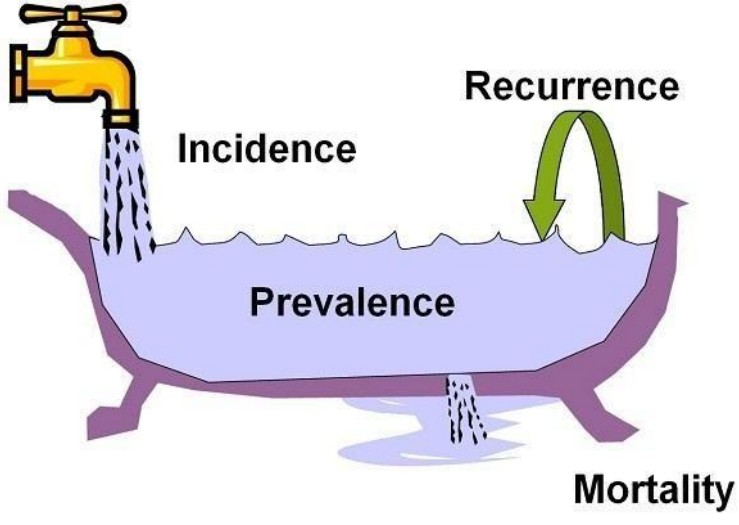
Epidemiological measures

- **Prevalence** – No. of existing cases
  - No. cases / total population at risk
  - 22 / 200 = 0.11 = 11% = 110 per 1000 persons

- **Incidence rate** – No of new cases
  - No. new cases within a time period / total population at risk
  - 8 / 400 (200*2) = 0.02 = 2% = 20 per 1000 person years
  - 8 / 360 (yr 1 =200 yr 2 = 160) = 0.0222 = 2.2% = 22 per 1000 p/yrs
  - 8 / 200 (200 health at start year 1) = 0.04 = 4% = 40 per 1000 p/yrs

Epidemiological measures

Prevalence = Incidence*Disease Duration



*Over the last 20 years a number of new drugs have been used to lower the risk of death during and after acute myocardial infarction (heart attack). However, there has **not** been an increase in the prevalence of coronary heart disease (CHD) in the population. Why?*

Binary outcomes

- **Risk** – the risk of an event is the probability of the event occurring.
- Probabilities lie between 0 and 1 (0=impossible, 1=certain to occur)
- **Risk difference**

  $= Risk_{Exposed\_group} - Risk_{Control\_group}$

  = 2% - 0.5% = 1.5%=0.015
- **Risk ratio/relative risk**

  $= Risk_{Exposed\_group} / Risk_{Control\_group}$

  = 2% / 0.5% = 4
- **Odds ratio -** is the ratio of an event occurring vs not occurring

  $= Odds_{Exposed\_group} / Odds_{Control\_group}$

  = 4/196 / 1/199 = 4.001

Binary outcomes

Number Needed to Harm (NNH) = 1 / Risk Difference

$$Risk\ Difference = Risk_{Exposed} - Risk_{Control}$$

NNH gives the average number of patients that need to be exposed in order for one additional ill-health event to occur

Number Needed to Treat (NNT) = 1 / Risk Difference

$$Risk\ Difference = Risk_{Treated} - Risk_{Control}$$

NNT gives the average number of patients that need to be treated in order for one additional ill-health event to occur

Note: The risk difference in the calculation of the NNT should be expressed as a number between 0 and 1, NOT as a percentage.

Example question

A cohort study evaluated the relationship between dietary calcium supplementation and hip fractures in post-menopausal women. 100 women took calcium supplements and 100 women took placebo tablets.

Over the three year period, five women had hip fractures in the calcium group and ten women had hip fractures in the placebo group. The 95% confidence interval is 0.18 to 1.4.

What is the RISK of a hip fracture in the TREATED group? Select ONE option only.

A. 0.01
B. 0.05
C. 0.1
D. 0.5
E. 1.0

Probability of event in the calcium group: 5/100=0.05

Example question

A cohort study evaluated the relationship between dietary calcium supplementation and hip fractures in post-menopausal women. 100 women took calcium supplements and 100 women took placebo tablets.

Over the three year period, five women had hip fractures in the calcium group and ten women had hip fractures in the placebo group. The 95% confidence interval is 0.18 to 1.4.

What is the RISK RATIO? Select ONE option only.

A. 0.01
B. 0.05
C. 0.1
D. 0.5
E. 1.0

Probability of event in the placebo group: 10/100=0.1

Risk ratio=0.05/0.1=0.5

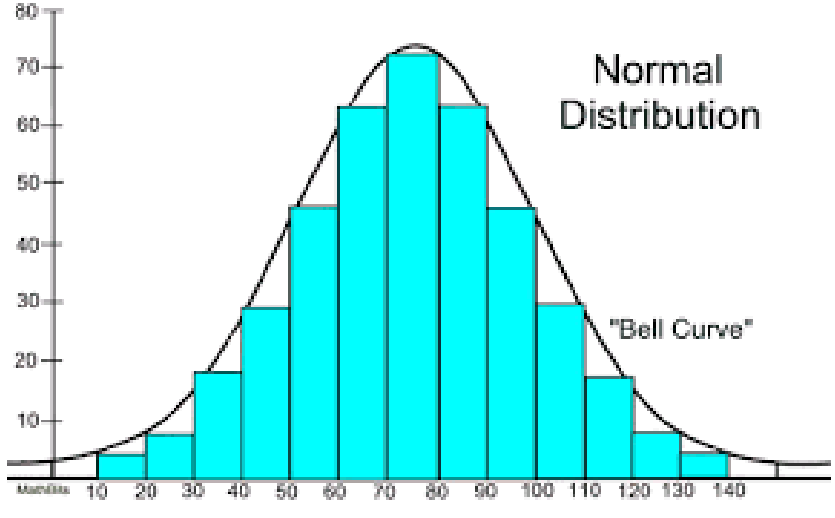
Interpretation: the event is twice as likely to occur in the placebo group than the treated group.

Types of Data/Outcome

- **Binary** – yes/no e.g. death, onset of type II diabetes
- **Categorical** – distinct groups ordinal/nominal
- **Count** – number of events occurring – a discrete number ≥0 (0, 1, 2, 3,….)
- **Continuous** – any value along a scale – height, age, percentage.

Continuous outcomes

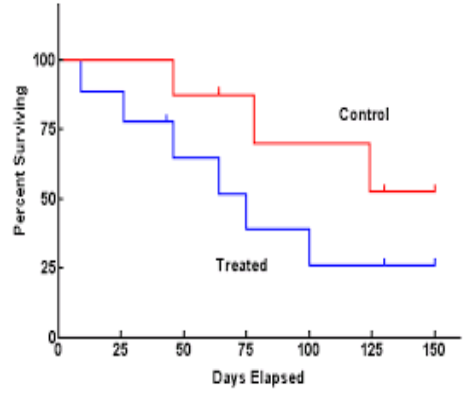
- A continuous outcome measure is typically summarised by the **mean**.
- The comparison of a continuous outcome between treated and non-treated or exposure and non-exposure groups is typically through a comparison of means.



Survival outcomes

- **Survival outcomes** – analysed differently to binary outcomes to account for the fact that 1) people take different lengths of time for the event to occur and 2) some people might go on to observe the event after the study has finished.
- **Hazard**: the probability of the event occurring in the next moment
- **Hazard ratio**

  = $\text{Hazard}_{\text{Treatment\_group}}$ / $\text{Hazard}_{\text{Control\_group}}$



Factoring in uncertainty

- The point estimates of an effect (e.g. relative risk, hazard ratio) is our ‘best guess’ of the true parameter value based on our data.
- Point estimates need to be accompanied by a measure of uncertainty to reflect the fact that we have estimated the point estimate from a sample.
- **Standard error** - a measure of the statistical accuracy of a point estimate. The larger the standard error, the larger the degree of uncertainty.
- **Confidence interval** – a range of plausible values for the true parameter of interest based on the data collected.

Confidence intervals

- 95% confidence intervals are most commonly presented.
- If we were to take multiple samples and estimate the point estimate in each separate sample, 95% of the 95% confidence intervals would contain the true parameter value.
- If the confidence interval contains the ‘null effect’ (1 for a ratio, 0 for a difference), we have no evidence to suggest that there is a difference between the two groups.

Example question

A cohort study evaluated the relationship between dietary calcium supplementation and hip fractures in post-menopausal women. 100 women took calcium supplements and 100 women took placebo tablets.

Over the three year period, five women had hip fractures in the calcium group and ten women had hip fractures in the placebo group. **The 95% confidence interval is 0.18 to 1.4.**

What is the RISK RATIO? Select ONE option only.

A. 0.01
B. 0.05
C. 0.1
D. 0.5
E. 1.0

Probability of event in the placebo group: 10/100=0.1

Risk ratio=0.05/0.1=0.5

Interpretation: the event is twice as likely to occur in the placebo group than the treated group.

Hypothesis testing

- Hypothesis tests are widely used in medical research and are based on specifying a hypothesis that states there is no difference in the outcome between two or more groups.
- The data are then examined to see if they are consistent with that hypothesis.
- **Null hypothesis**: the risk ratio comparing the risk of hip fracture in women who did and did not receive calcium supplementation is equal to 1.
- **Null hypothesis**: the risk ratio comparing the risk of hip fracture in women who did and did not receive calcium supplementation is the same in women who smoke and women who do not smoke.

Hypothesis testing

- The choice of statistical test will depend on the data.

	Plausibly Continuous and Normal	**Ordinal or Ordered Categorical**	**Binary and Unordered Categories**
Comparison of Independent Two Groups	Box-plot Independent groups t-test	Box-plot or Cross-tabulation of ordered categories Mann-Whitney U-test	Cross-tabulation Chi-squared test Fisher's exact test
Comparison of more than Two groups	Analysis of variance (ANOVA)	*Kruskal Wallis analysis of Variance**	Cross-tabulation Chi-squared test
Comparison of two related outcomes	Paired samples t-test	Wilcoxon Matched Pairs	McNemar's Test
Relationship between a dependent variable and one or more independent variables	Scatter plot Regression *Pearson's correlation coefficient*	*Spearman correlation or Kendall's correlation coefficient*	*Phi coefficient* *Logistic Regression*

p-values

- The output of most statistical tests is a ***p*-value**.
- A *p*-value is defined as ‘the probability that we observed a the result at least as extreme as the one we did given that the null hypothesis is true’.
- The smaller the *p*-value, the more unlikely it is that we would have collected the data we did given that the null hypothesis is true…this means that the smaller the *p*-value, the more likely it is the null hypothesis is not true.
- A popular cut-off for determining whether there is sufficient evidence against the null hypothesis is ***p*<0.05**.

Type I and Type II errors

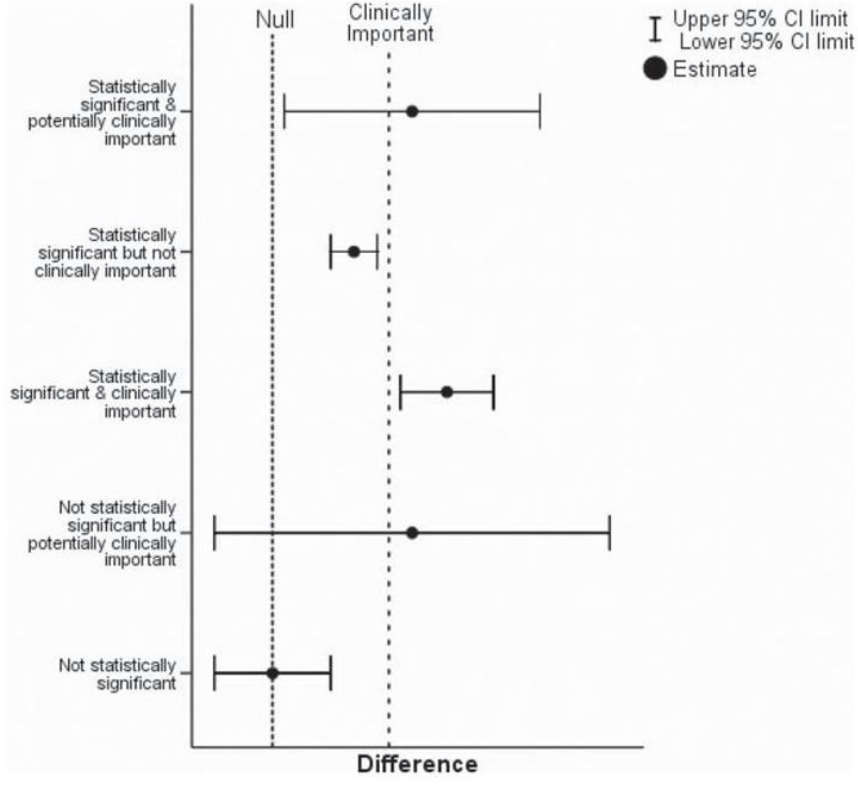
		True situation	
		Null hyp. is true	Null hyp. is false
Researchers decision	Reject Null hyp.	Type I error	Correct
	Fail to reject Null hyp.	Correct	Type II error

- **Type I error**: the probability of rejecting the null hypothesis when it is in fact true.
- **Type II error**: the probability of failing to reject the null hypothesis when it is in fact false.

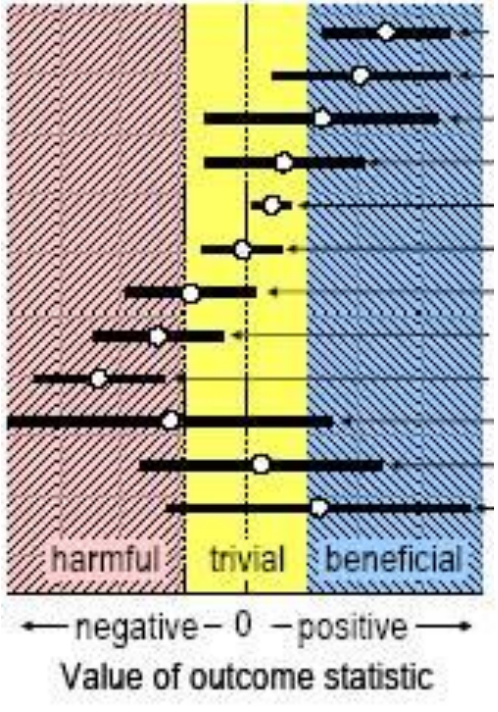
Statistical power and sample size

- The **power** of a statistical test is the probability of rejecting the null hypothesis given that it is false.
- Studies generally use a power of either 80% or 90%.
- The power of a study can be increased by increasing sample size.
- The larger the sample size, the better!
- Studies with small samples will tend to have wide confidence intervals which reduces our confidence in the results.

Statistical significance vs clinical significance



Interpreting effect sizes

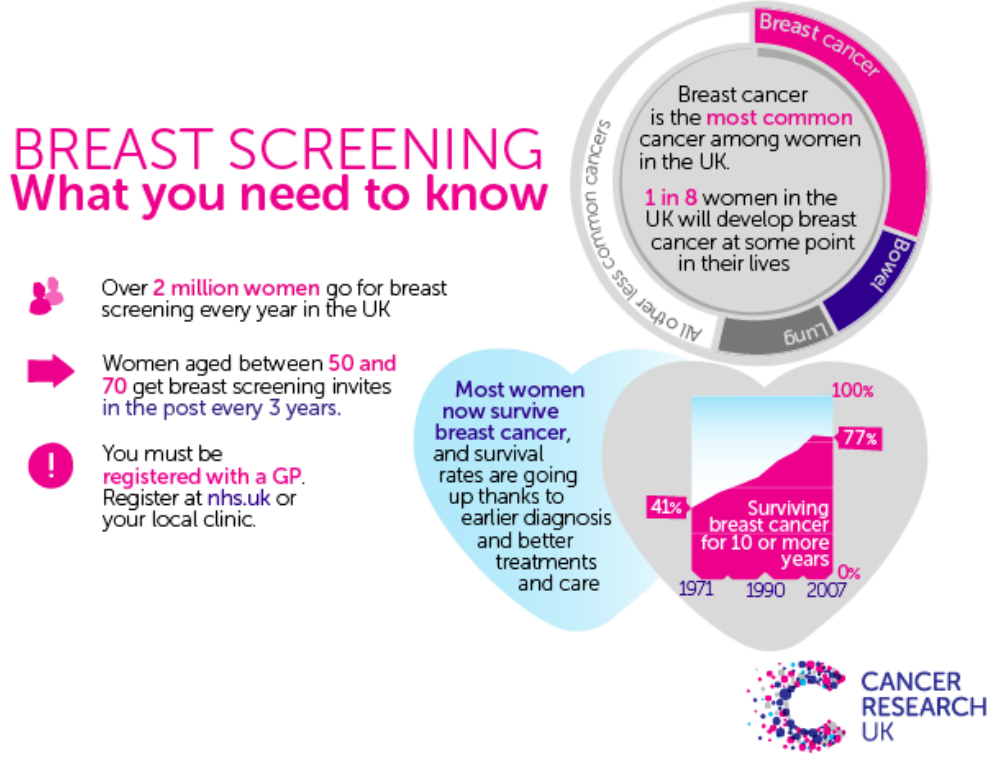


Almost certainly beneficial
Probably beneficial
Possibly beneficial
Possibly trivial
Very likely trivial
Probably trivial
Possibly trivial
Likely harmful
Very likely harmful
Unclear
Unclear
Unclear

Exercise

Screening

Example



http://www.cancerresearchuk.org/about-cancer/type/breast-cancer/about/screening/infographic

Definition and purpose of screening

- Screening is used to categorise seemingly healthy individuals into one of several risk or disease categories.
- Screening involves the use of a diagnostic test.
- A diagnostic test study compares a new method of measurement to a true ‘gold standard’ method of measurement.

Sensitivity and specificity

- A diagnostic test will always results in some false positive and false negative results.
- E.g. a **false positive** result would occur when the test indicates a women has early stage breast cancer when she does not.
- E.g. a **false negative** result would occur when the test indicates a women does not have early stage breast cancer when she does.

Sensitivity and specificity

- The **sensitivity** of a diagnostic test is the proportion of truly abnormal persons who are correctly identified as abnormal by the test.

How good is this test at picking up breast cancer in women who have this condition?

- The **specificity** of a diagnostic test is the proportion of truly normal persons who are correctly identified as normal by the test.

How good is this test at ruling out breast cancer in women who don't have this condition?

- We want to minimise both!

Sensitivity and specificity

Sensitivity = $\dfrac{\text{No. True Positives}}{\text{No. True Positives + No. False Negatives}}$

$\dfrac{\text{No. True Positives}}{\text{Total No. sick individuals in population}}$

Specificity = $\dfrac{\text{No. True Negatives}}{\text{No. True Negatives + No. False Positives}}$

$\dfrac{\text{No. True Positives}}{\text{Total No. healthy individuals in population}}$

Sensitivity and specificity

Screening	Present	Absent	Total
Positive	127	497	624
Negative	63	19313	19376
Total	190	19810	20000

Sensitivity = 127/190 = **67%**

Specificity = 19313/19810 = **98%**

Positive predictive value

- The **positive predictive value (PPV)** of a diagnostic test is proportion of persons with abnormal test results who are abnormal.

$$\textbf{PPV} = \frac{\text{No. True Positives}}{\text{No. True Positives} + \text{No. False Positives}}$$

Screening	Present	Absent	Total
Positive	127	497	624
Negative	63	19313	19376
Total	190	19810	20000

Positive Predictive Value = 127/624 = **20%**

Additional info

Features of well designed study

- A **well designed study** should have the following features.
  - A well defined, pre-specified aim and clinical question.
  - A discussion of sample size.
  - Generalisability.
  - Transparency.
  - A thorough discussion of the limitations of the data.
  - A thorough discussion of the limitations of the methodology (e.g. possible types of bias).

- **Publication bias** occurs when the outcome of an experiment or research study influences the decision whether to publish or otherwise distribute it.

Questions?