Epidemiology and statistics
Session 10: Revision

Robert Froud

Session one concepts

- Journals
- Risk
- Incidence
- Prevalence
- Mortality rates
- Age-specific rates
- Person-years
- Hazard ratios

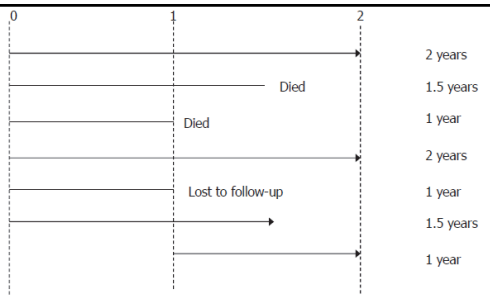
Mortality rate per 100,000

= Number of deaths/(person years at risk) x 100,000

For malaria example
Malaria deaths (0 to 69 years) = 18 per 100,000

Age band	Deaths per 100,000
1 – 59 months	44
5-14 years	12
15 to 29 years	8
30-44 years	10
45-59 years	27
60-69 years	75

Barts and The London School of Medicine and Dentistry www.smd.qmul.ac.uk



Risk = d / N = 2 / 7 = 0.29

Rate = d / Σt_i = 2 / (2+1.5+1+2+1+1.5+1) = 2/10
= 0.2 deaths/person-year

One two three four five six seven eight nine

Hazard ratios

- Those eating little meat
  - Incidence rate (Cancer/CVD) =1713/151212 x 100,000
    =1133 per 100,000
- Those eating most meat
  - Incidence rate =2130/151315 x100,000
    =1408 per 100,000

- Ratio of rates=1408/1133=1.25

- Hazard is probability of event at set time point
- Hazard ratios can be interpreted as ratio of incidence rates.
- Probability of dying is 1.24 times higher in those eating most meat compared with those eating least
- Cox proportional hazards model

Barts and The London School of Medicine and Dentistry www.smd.qmul.ac.uk

One two three four five six seven eight nine

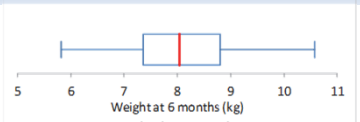
Session two concepts

- Types of data (e.g Numerical, categorical & sub-categories)
- Central tendency (mean, median, mode)
- Spread (Centiles, SD, variance)
- Graphs & plots (e.g. box, scatter, histograms)
- Probabilities
- Proportions
- Normal distribution
- Reference groups
- PICO(s)

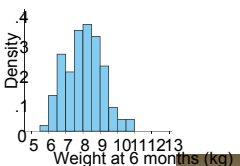
Barts and The London School of Medicine and Dentistry www.smd.qmul.ac.uk

One two three four five six seven eight nine

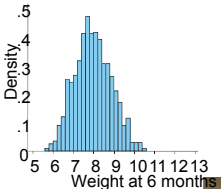
Spread, central tendency, and normal distribution



Box and whisker plot of of weight of 100 boys



Histogram of weight of 100 boys



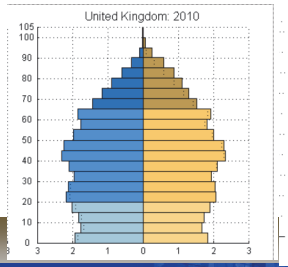
Histogram of weight of 1000 boys

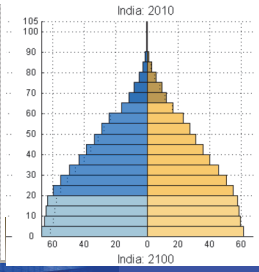
Barts and The London School of Medicine and Dentistry

One two three four five six seven eight nine

Session three concepts

- Populations & Data sources (Birth & death registration, HES, GP data)
- Tables
- Standardisation
- Population pyramids





Barts and The London School of Medicine and Dentistry

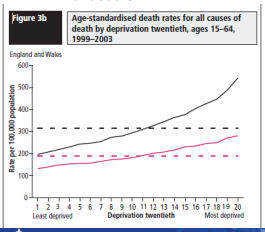
One two three four five six seven eight nine

Taking age into account

Direct standardisation
Uses rates from population of interest and apply to standard population

Indirect standardisation
Uses rates from reference population and calculates expected number of deaths

7000x0.3/1000=2.1



Males UK 2010

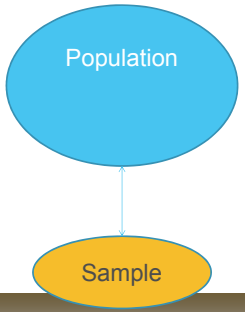
Age	Death rate/ 1000	Number in Standard population	Expected number of deaths
15–19	0.3	7000	2.1
20–24	0.5	7000	3.5
25–29	0.6	7000	4.2
30–34	0.8	7000	5.6
35–39	1.2	7000	8.4
40–44	1.7	7000	11.9
45–49	2.5	7000	17.5
50–54	3.9	7000	27.3
55–59	6.2	6000	37.2
60–64	9.7	5000	48.5

Barts and The London School of Medicine and Dentistry

One two three four five six seven eight nine

Session four concepts

- Proportions
- ARR, RR, OR
- 2x2 tables
- Confidence intervals
- P-values
- Chi-squared test
- Probability
- Inference
- Erikson framework
- (Sensitivity and specificity)



Barts and The London School of Medicine and Dentistry

www.smd.qmul.ac.uk

One two three four five six seven eight nine

Confidence intervals and P-values

If 95% CI:

95% probability that the interval spans the true population parameter
(assuming adequate internal validity)

P-value is the **probability of observing this value, or a more extreme value, if the null-hypothesis is true**
(again assuming adequate internal validity)

One two three four five six seven eight nine

2x2, OR, RR, ARD

	Whooping cough	Not whooping cough	TOTALS
Immunised	55	104	159
Not immunised	9	4	13
TOTALS	64	108	172

Odds ratio = ratio of two odds
= (55/104) / (9/4)
= (55 X 4) / (9 X 104)
= 0.23 = odds of getting whooping cough if immunised compared to not being immunised

RR? ARD?

One two three four five six seven eight nine

Session five concepts

- Probability
- Incidence
- Prevalence
- Sensitivity & specificity
- Positive predictive value
- Lead time and lead time bias
- Wilson Junger criteria

Diagnosis using serology

Test says they do = 76%

Sensitivity

People who really **do** have whooping cough

Test says they don't = 99%

Specificity

People who really **don't** have whooping cough

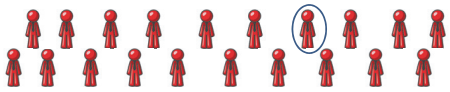
Two by two table

Likelihood of cancer if positive test = positive predictive value =180000/3510000=5%

	Breast cancer	Not breast cancer	TOTALS
Positive test	180,000	3,330,000	3,510,000
Negative test	20,000	29,970,000	29,990,000
TOTALS	200,000	33,300,000	33,500,000

Likelihood of not having cancer if negative test = negative predictive value =29970000/29990000=99.9%

Positive and negative predictive value



If I have a positive test what is the likelihood that I have breast cancer? Positive predictive value = 1/20 = 5%

If I have a negative test what is the likelihood that I don't have breast cancer? Negative predictive value is over 99/100 = 99%

Should we screen?

- Screening detects disease earlier
- How do you measure increase in life expectancy?

- **Lead time** is the length of time between the detection of a disease (usually based on new, experimental criteria) and its usual clinical presentation and diagnosis (based on traditional criteria)
- **Lead time bias** is the bias arising in survival analyses when the disease is detected earlier in one group than the other

Wilson and Jungner criteria (WHO 1968)

1. The condition should be an important health problem.
2. There should be a treatment for the condition.
3. Facilities for diagnosis and treatment should be available.
4. There should be a latent stage of the disease.
5. There should be a test or examination for the condition.
6. The test should be acceptable to the population.
7. The natural history of the disease should be adequately understood.
8. There should be an agreed policy on whom to treat.
9. The total cost of finding a case should be economically balanced in relation to medical expenditure as a whole.
10. Case-finding should be a continuous process, not just a "once and for all" project.

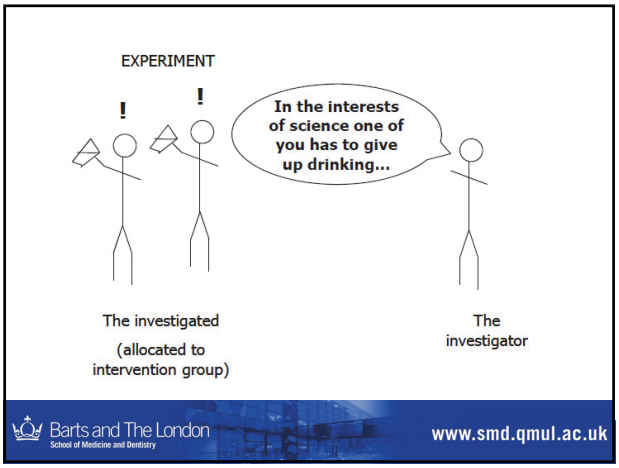
One two three four five six seven eight nine

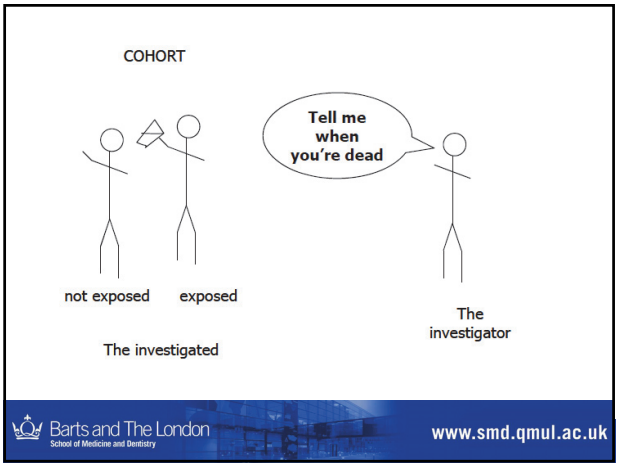
Session six concepts

- Study types
- Incidence
- Odds ratios
- Chi-squared test
- Cohort study
- Case-control study

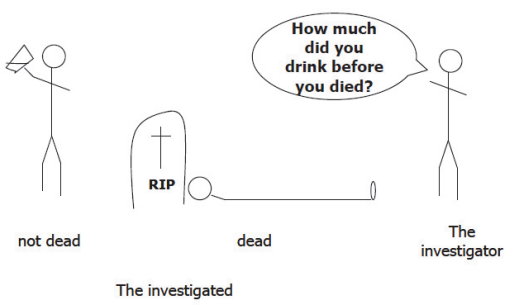
Ecological studies

- Look at correlations between exposure and outcome
  - Geographical (within or between countries)
  - Over time
- Collect published data/routine statistics on:
  - Risk factors eg national food consumption data
  - Disease eg mortality rates, published survey data
- Compare characteristics of *populations* (not individuals)

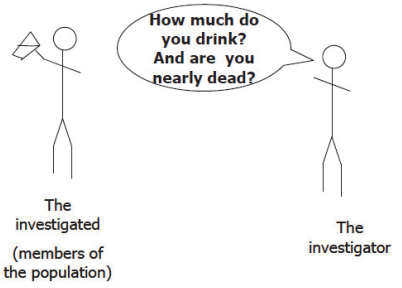




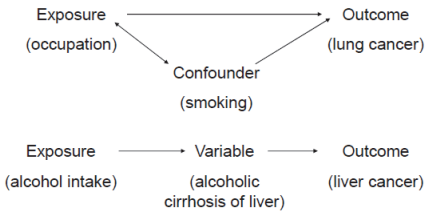
CASE-CONTROL

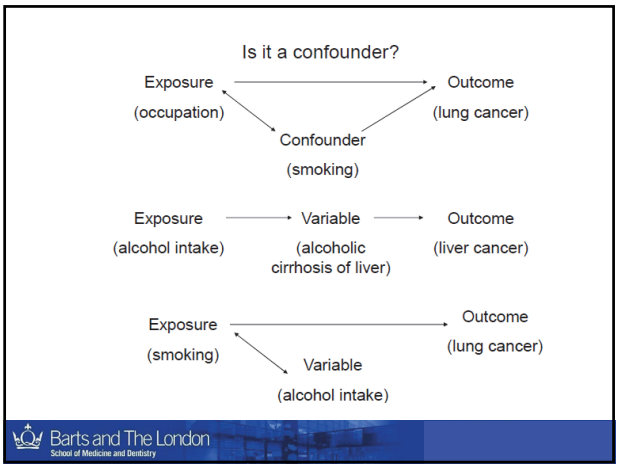


CROSS SECTIONAL



Is it a confounder?

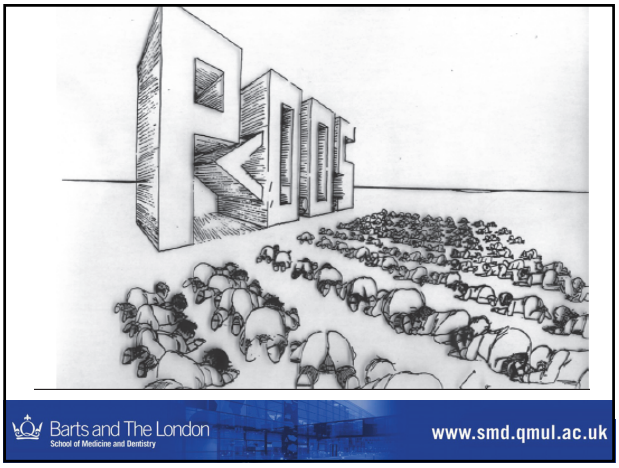




Session seven concepts

- Cohort and historical cohort studies
- Types of data (numerical, categorical)
- Confidence intervals
- P-values and statistical significance
- Clinical importance
- Chi-squared test
- Relative risk (and comparison to ORs)
- Trials
- Confounding
- Coding
- Bradford Hill criteria

Barts and The London
School of Medicine and Dentistry

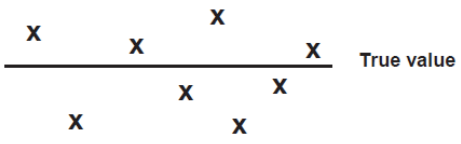


Can we believe our results?
First we need to consider:

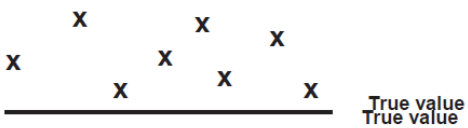
- *Chance*
- *Bias*
- *Confounding*

⇨ Is it causal?

Random error



Bias



Causal inference in observational studies

- Bradford Hill "criteria"
  - Size of effect
  - Dose response
  - Consistency
  - Biological plausibility
  - Temporality

Bradford Hill (1965)

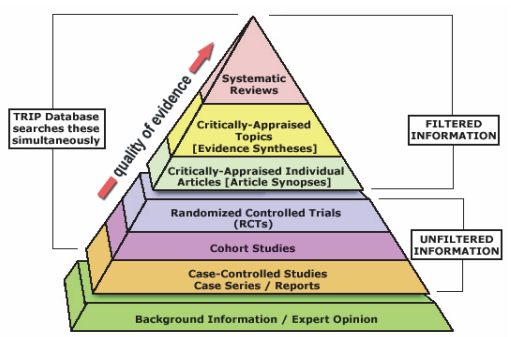
- **Temporality**
- **Strength**
- **Dose-response relationship (trend)**
- **Reversibility**
- Biological plausibility
- Consistency
- Specificity
- Analogy
- Coherence

(Proc R Soc Med 1965; 58: 295-300)

Session eight concepts

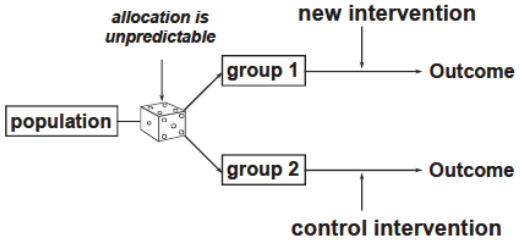
- Hierarchy of evidence
- Trials
- Randomisation
- Blinding
- Confidence intervals
- Statistical significance
- P-values
- Chi-squared test
- Logistic regression
- Confounding
- ITT & Per protocol
- Clinical importance vs statistical significance
- Ethics

Hierarchy of evidence



Barts and The London School of Medicine and Dentistry

Randomised controlled trial



Barts and The London School of Medicine and Dentistry www.smd.qmul.ac.uk

Randomisation

Block randomisation – *e.g.* blocks of 6:

A
B
B
A
A
B

Each block of 6 contains exactly
3 A's & 3 B's, but randomly arranged

A
B
A
B
B
A

Block randomisation ensures roughly equal numbers in each group

Barts and The London School of Medicine and Dentistry www.smd.qmul.ac.uk

One two three four five six seven eight nine

Session nine concepts

- PICO(s)
- Systematic review
- Meta analysis
- Forest plots
- Odds ratios
- Relative risk
- Confidence intervals
- Statistical significance
- P-values
- Confounding
- Descriptive statistics
- Risk difference
- NNT
- t-test
- Regression

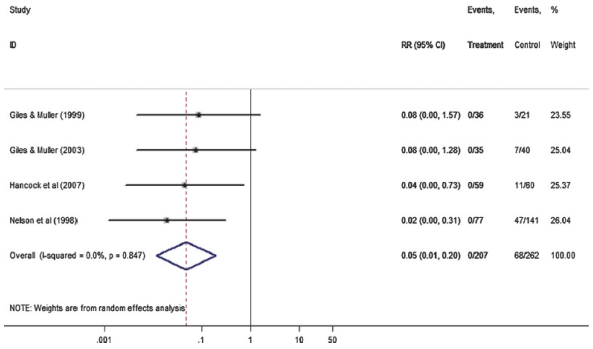
Describing the Research Question: PICO

P	Population	Where and how are participants recruited? What are the inclusion and exclusion criteria?
I	Intervention	What is the active intervention? It should be described in enough detail to allow another researcher to replicate it
C	Comparison	What is the control? This might be an established treatment or 'routine care', or else a placebo, sham, or inactive control
O	Outcome	How and when will outcome be assessed?

Systematic reviews

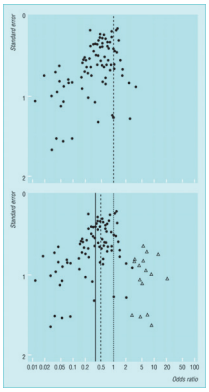
- Develop research question
- Scoping
- Refine research question
- Search abstracts ->reject/accept
- Search full-texts -> reject/accept
- Abstract data (independent – team up)
- Analyse (*e.g.* meta-analysis, or descriptive)
- Write-up dissertation
- Publish results

Manual therapy vs drugs



Asymmetrical funnel plot of 89 randomised controlled trials comparing homoeopathic medicine with placebo
Combined odds ratio 0.41

Same funnel plot with 'missing' studies added
Combined odds ratio 0.52



Sterne, J. A C et al. BMJ 2001;323:101-105

BMJ

41

Difference between proportions

Does tamiflu (oseltamivir) prevent respiratory tract infections requiring antibiotics in children with confirmed influenza

28% in those given placebo
17% in those given tamiflu
Difference 11 (4 to 19)

Barts and The London School of Medicine and Dentistry

www.smd.qmul.ac.uk

Number needed to treat

Number of people you need to treat to avoid one adverse event

NNT = 1/Risk difference

For tamiflu

Risk difference =0.11

Number needed to treat = 1/0.11=9

95% confidence interval (5 to 29)

Continuous outcomes

Mean duration of symptoms

One study by Vrese

Control mean(SD) = 8.9 (1.0) days

Probiotics mean(SD) = 7.0 (0.5) days

Difference = 1.9

95% confidence interval 1.75 to 2.05; *P* <0.05

Continuous outcomes

Simple comparison of two groups

Use a t-test

Gives confidence interval for difference

Adjusting for confounders

Use regression model

Present results as “adjusted difference”