Epidemiological methods

Dr Jitka Pikhartova
Department of Epidemiology and Public Health
University College London

Objectives:

At the end of the session students should be able to:

- Differentiate between different types of data
- Describe the structure of an epidemiological dataset
- Define and calculate measures of disease occurrence and measures of association
- Describe the basic features of the main types of epidemiological studies
- Explain the main features of bias, confounding, chance
- Be familiar with causation criteria

Introduction

- Until 1950s, the term "epidemiology" was mainly used for studies of communicable diseases
- Later, it was suggested that a new field of study should be created to look at non-epidemic diseases
- The meaning of "epidemiology" was broadened to cover also non-communicable diseases

Definition of epidemiology

- A modern definition of epidemiology is thus very general:

- **Epidemiology is the study of the distribution and determinants of disease in population**

Much of epidemiological research is taken up trying…

- to establish associations between exposures and disease rates
- to measure the extent to which risk changes as the level of exposure changes
- to establish whether the associations observed may be truly causal (rather than being just consequence of bias or chance)

- Epidemiology has a major role in developing appropriate strategies to improve public health through prevention
  - public health has wider meaning in this sense; it is about the health of the whole population.
  - it does not cover only classic areas, such as immunization or monitoring of diseases, it also covers factors such as poverty, smoking, nutrition
- In this sense, epidemiology has a crucial role in trying to put into perspective the effects on population health of different risk factors.

Epidemiology

- The study of the **distribution** and **determinants** of the **frequency** of health-related outcomes in specified populations
- Quantitative discipline
- Measurement of disease / condition / risk factor frequency is central to epidemiology

7

Variables (outcomes/risk factors)

- Binary
  - Deaths (y/n)
  - Sex (m/f)
- Categorical (ordinal or nominal)
  - Frequency of drinking (never, 1-3 times a month, 1-3 times a week, 4 times a week or more often)
  - Severity of pain (none, some, a lot)
  - Marital status (single, married/in partnership, divorced, separated, widowed)
  - Country of birth (Czech R, Slovakia, Poland,Austria, Germany, Ukraine, Hungary)
- Continuous
  - BMI, blood pressure, etc.

8

What type of variable is…

- Self-rated health (Very poor, poor, average, good, very good)
- Total cholesterol concentration
- Economic activity (employed, unemployed, housewife, pensioner)
- Risk of CVD death in the next 10 years (score)
- Having lung cancer or not
- Quartile of income
- Sex
- Social class (upper, upper-middle, middle, working, lower)

9

What type of variable is…

- Self-rated health =Categorical (ordinal)
- Total cholesterol concentration = Continuous
- Economic activity = Categorical (nominal)
- Risk of CVD death in the next 10 years (score) = Continuous
- Having lung cancer or not = Categorical (binomial)
- Quartile of income = Categorical (ordinal)
- Sex = Categorical (binomial)
- Social class = Categorical (ordinal)

10

Binary outcomes: "cases" vs. "non-cases"

- Persons with disease = "cases"
- Definition of case is crucial
- E.g.
  - Obesity: BMI≥30
  - Hypertension: SBP≥140 mm Hg or DBP≥90 mm Hg or treatment
  - High cholesterol: ≥6.2 mmol/L
- Can be complex in clinical settings
  (e.g. metabolic syndrome, depression, etc.)
- **But must always be clearly specified**

11

Measures of disease frequency

- Used for binary outcomes
- Require a numerator and denominator

$$= \frac{\text{number of persons with disease}}{\text{number of persons examined}}$$

- expressed as X per 1,000 persons (or per 100,000 etc.)

12

Numerators and denominators

Example:

- The number of cancer cases in the UK is 247,667 whereas in Belgium it is 47,948
- The UK has a bigger problem in numerical terms
- But do Belgians have lower risk of getting cancer?
  - Numerators alone are meaningless
  - We need both **numerators** AND **denominators**

13

Numerators and denominators

- The number of cancer cases in the UK is 247,667 whereas in Belgium it is 47,948
- UK: 247 667 / 65 000 000 = 0.00381 = 381 per 100 000
- Belgium: 47 948 / 11 000 000 = 0.00436 = 436 per 100 000

14

What we measure…..

a) Prevalence

= number of **existing** cases
population of interest at a defined time

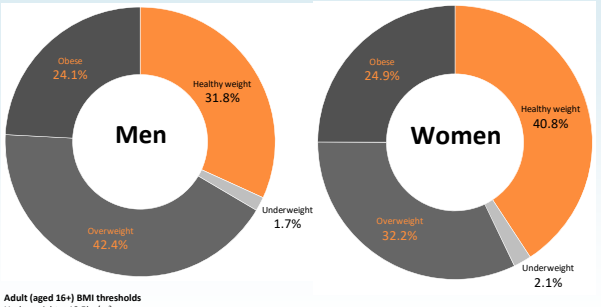
- Unable to work now for health reasons
- Occupational injury ever
- Ever wheezing or whistling in the chest
- Hangover in the last 12 months
- Headache today

NOTE a **denominator** is needed for prevalence

15

Adult prevalence by BMI status

Health Survey for England (2008-2010 average)



Adult (aged 16+) BMI thresholds
Underweight: <18.5kg/m²
Healthy weight: 18.5 to <25kg/m²
Overweight: 25 to <30kg/m²
Obese: ≥30kg/m²

16

b) Incidence

= number of **new** cases in a given time period
total population at risk

17

Exercise

- In 2014, **55,222** new cases of breast cancer were diagnosed in the UK.
- Approximately **65M** people in the UK
- Most cases in women (only **389** cases in men)
- Population at risk?
- Cumulative incidence of breast cancer in the UK in 2014 in females was ?

???

???

18

- In 2014, **55,222** new cases of breast cancer were diagnosed in the UK.
- Approximately **65M** people in the UK
- Most cases in women (only **389** cases in men)
- Population at risk?
- Cumulative incidence of breast cancer in the UK in 2014 in females was ?

$$\frac{55{,}222-389}{65M/2} = \frac{54{,}833}{32.5M} = 0.001687 = 168.7/100{,}000$$

19

Example

3-year study with a sample size of 100, outcome of interest was fatal heart disease.

	year 1	*year 2*	*Study ends*
Developed outcome	6	5	4
Dropped out	4	10	-
Sample at risk	90	75	-

- 10 participants were followed for 1 year
- 15 participants were followed for 2 years
- 75 participants were followed for 3 years

Total person-years:

Incidence Rate:

20

3-year study with a sample size of 100, outcome of interest was fatal heart disease.

	year 1	*year 2*	*Study ends*
Developed outcome	6	5	4
Dropped out	4	10	-
Sample at risk	90	75	-

- 10 participants were followed for 1 year
- 15 participants were followed for 2 years
- 75 participants were followed for 3 years

Total person-years of follow up = (10x1) + (15x2) + (75x3) = 265 person-years at risk

Incidence rate = 15 / 265 = 0.057 = 5.7 cases per 100 person-years

21

Relationship between prevalence and incidence

- The prevalence of a health-related outcome depends both on the incidence rate and the time between onset and recovery or death
- **Prevalence = Incidence x Average disease duration**

22

Exercise

- Population of 10,000 people
- 10 new cases of cancer a year
- 20 registered cases at any time
- Average duration of (survival from) the cancer is…

23

Exercise

- Population of 10,000 people
- 10 new cases of cancer a year
- 20 registered cases at any time
- Average duration of (survival from) the cancer is…
  20/10 (prevalence/incidence) = 2 years

24

c) Mortality

- **= number of deaths / total population**
- Rate (or risk)
- = $\frac{\text{the number of deaths in a specified population}}{\text{the number of that population}}$ **/per unit time**
- If the mortality rate is to be calculated in a given year, the mid-year population is usually used as the denominator
- Mortality rate is always expressed as deaths per X (e.g. 1,000 persons per year)

25

Example

– A city has a population of 900,000;
– 30,000 deaths occur in a 3-year period

– Mortality rate for the period = $\frac{30\,000}{900\,000}$
= 0.0033 or 33 deaths per 1,000
= 11 deaths per 1,000 per year

26

All mortality rates MUST include

- Number of deaths = *numerator*
- Population size (in which these deaths were counted) = *denominator*
- Time period (during which these deaths happened)

27

Exercise

Which piece of information were necessary to calculate following result:

1.5 deaths/10,000 population per day

1. All deaths occurred in hospital
2. The population is 29,661
3. 53% of deaths were males
4. The deaths occurred over 3 months
5. 404 deaths happened
6. Tuberculosis caused 17% of the death**s**

28

cases/ population per number of days ….recalculated to 10,000

- The number of deaths = 5.
- The population size = 2.
- The time period in which the deaths occurred = 4.

29

Which piece of information were necessary to calculate following result:

1.5 deaths/10,000 population per day

cases/ population per number of days ….recalculated to 10,000

1. All deaths occurred in hospital
2. The population is 29,661
3. 53% of deaths were males
4. The deaths occurred over 3 months
5. 404 deaths happened
6. Tuberculosis caused 17% of the death**s**

30

Mortality rates

- **All-cause mortality rates**
- **Cause-specific mortality rate**
- **Crude mortality rates**
- **Standardized mortality rate**

31

- **All-cause mortality (rates):** refers to the total number of deaths per 1,000 people per year

32

- **Cause-specific mortality rate**

= total number of deaths due to a specific cause
(population at risk x period of time)

33

- **Crude mortality rates:** no care has been taken for age structure of the population
  - Counts all deaths
    - All cases
    - All ages and sexes
  - Denominator includes entire population
    - All ages and sexes

34

- **Standardized mortality rate** refers to a mortality rate which is age-standardized in order to permit comparisons between different countries, regions etc.

=also **age-specific mortality rate**

  - Infant mortality rate
  - Maternal mortality rate
  - Under-5 mortality rate

35

Other commonly used measures

- **Perinatal mortality rate** is the number of neonatal and fetal deaths (stillbirths) per 1000 births
- **Case fatality rate** is the rate of death among people who already have a condition, usually in a defined period of time. usually measured as a decimal or as a percent.
- **Survival rate** is the proportion of people who remain alive for a given period of time after diagnosis of disease. E.g. breast cancer has 5-year survival rate around 70%.

36

Recently often measured and mentioned…

- **Number of excess deaths**

Hypothetical number of deaths caused by the emergency itself

Or

Number of deaths that would not have occurred had the emergency not happened

=Difference between pre-emergency mortality rate and mortality rate found during emergency **x** population size

37

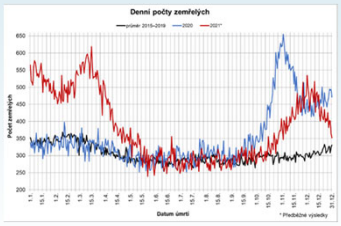
Most recent excessive deaths

- Covid

38

Most recent excess deaths

- Covid



Source: data from CSO

39

Exercise & a bit of history

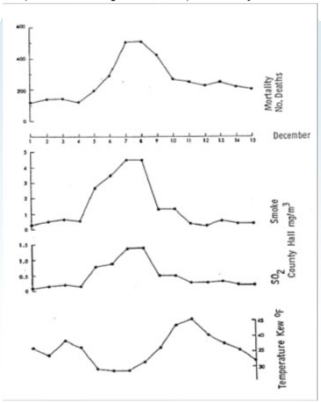
Smog in London

During the first half of **December of 1952**, the London area experienced periods of fog culminating in one of the most intense in memory lasting from the morning of Friday 5 December to early in the morning of Tuesday 9 December and then dispersed quickly when the weather changed.

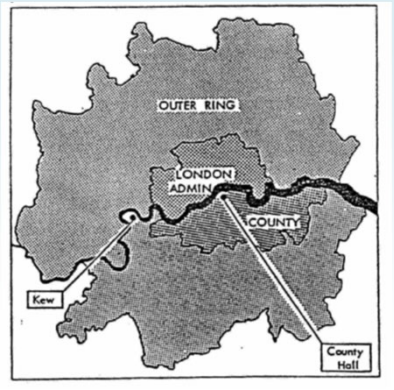
Air pollution and meteorological factors, particularly low temperatures, were suggested as possible causative or contributory agents. A period of cold weather, combined with an anticyclone and windless conditions, collected airborne pollutants - mostly arising from the use of coal - to form a thick layer of smog over the city.

40

Figure 1 Numbers of deaths in London AC, mean daily temperature at Kew and mean atmospheric pollution (smoke and sulphur dioxide) at County Hall between 1st and 15th December 1952.



41



42

Table 1 Number of deaths registered in Greater London and in the Great Towns outside London in the weeks ending 8 November 1952 to 10 January 1953

Week ending	Nov 8	Nov 15	Nov 22	Nov 29	Dec 6	Dec 13	Dec 20	Dec 27	Jan 3	Jan 10
London Administrative County (AC)	693	747	753	853	945	2484	1523	1029	1372	1216
Outer Ring (remainder of Greater London	900	818	946	1049	1117	2219	1615	1205	1605	1418
Greater London	1593	1565	1699	1902	2062	4703	3138	2234	2977	2634
160 Great Towns minus London	3310	3410	3603	4140	4585	4749	4541	4238	4865	4903

Table 2 Deaths in Greater London by date of occurrence, December 1-15 1952, with daily meteorological data for this period

December	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Deaths															
London AC	112	140	143	120	196	294	513	518	430	274	255	236	258	220	213
Outer ring	147	161	178	168	210	287	381	392	362	269	273	248	245	227	212
Greater London	259	301	321	288	406	581	894	910	792	543	528	484	501	449	425
Temperature (°F)															
Daily mean (Kew)	36.9	34.2	39.0	36.5	29.5	28.9	28.9	31.5	36.0	43.3	45.1	40.1	37.2	35.2	32.0
Departure from av of 80 years	-5.2	-7.7	-3.5	-5.4	-12.1	-12.8	-12.3	-10.0	-1.5	+2.7	+5.0	-0.1	-4.2	-6.0	-8.8
Atmospheric pollution (mg/m³) (smoke)															
Mean (Kew)	0.34	0.34	0.19	0.42	1.47	1.73	0.87	1.19	0.47	0.17	0.19	0.24	0.32	0.29	0.18
Mean (County Hall)	0.30	0.49	0.61	0.49	2.64	3.45	4.46	4.46	1.22	1.22	0.32	0.29	0.50	0.32	0.32
Sulphur dioxide															
Parts/m;	0.09	0.16	0.22	0.14	0.75	0.86	1.34	1.34	0.47	0.47	0.22	0.23	0.26	0.16	0.16

43

Q1. Comment on findings in Table 1

a) This is 'numerator data' referring to cases and not to population-based measurements such as incidence or prevalence. However, over the short period involved, the assumption of a constant population at risk is not unreasonable.
You should have thought to plot the data (roughly).

b) Data presented by date of registration, not date of death —possible delays over the Christmas period.

c) November forms the 'baseline' against which the 'epidemic' can be assessed.

d) Comparison with other great towns suggests an epidemic specific to London, and most acute in Central London (London AC).

44

Q2. Define the period of excess mortality in London and its relationship to the prevailing weather conditions

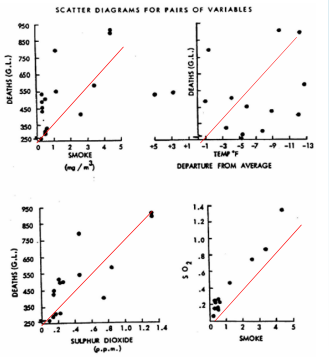
a) Compared to Dec 1-4, excess mortality peak at Dec 7-8 is 4-5 times baseline rate and has not entirely resolved by 15th.

b) Time course of smoke and SO_2 very similar to each other and closely followed by rise and fall in numbers of deaths. Latent period of 24-48 hours suggests an acute toxic mechanism.

c) Temperature changes less well matched by changes in numbers of deaths.

45

Figure 2 Scatter diagrams showing the relationship between number of deaths in Greater London between 1st and 15th December 1952 with air pollution and meteorological factors measured during that time, and a scatter diagram showing the association between smoke and sulphur dioxide in that period.



46

Q3. How might you proceed to further investigate the influence of fog on mortality?

a) Clarify the nature of the epidemic. Which causes of death were most affected? Which age groups? Were deaths confined to already sick?

b) Obtain more detail on the hazardous 'exposure'. Examine the time course and geographical distribution of fog, smoke and sulphur-dioxide. Compare the effect on outdoor and indoor workers?

c) Look for similar epidemics associated with fog elsewhere and in London in the past. (This was the first time that daily mortality returns had been examined — monthly figures would have shown a much less marked effect).

d) Follow time course of mortality after the fog to investigate delayed consequences (cause-specific data by age + sex most useful).

47

Now, we have data and we know basic measures

- Let's consider **2 groups** of individuals
- An **exposed** group (group with risk factor of interest) and **unexposed** group (without such factor of interest)
- We are interested in <u>comparing</u> the amount of disease (mortality or other health outcome) in the exposed group to that in the unexposed group

48

Measures of association

- Risk of disease, rate of disease in different groups of population
- Comparison of risks/rates

49

(Absolute) Risk

- Risk is the probability of new occurrence of disease among individuals in an initially disease-free population during a defined time period
- To calculate a risk (r), we divide the number of **new cases** (d) in the defined period by the **population** at risk at the beginning of the period (N);
  (d and N are referred to as the numerator and denominator, respectively)
  r = d / N over a defined period

50

- Risk is probability but is often multiplied by a suitable number (eg 100,000)

Example

- *In 1980, an annual risk of death was*
  14 per 1,000 in Kenya,
  10 per 1,000 in France
  26 per 1,000 in Malawi
  (United Nations, Demographic Yearbook)

51

Risk measures

- Risk in exposed (r_1)
- Risk in unexposed (r_0)

52

Risk ratio

$$RR = \frac{r1}{r0} = \frac{\text{incidence in the group with attribute/exposure}}{\text{incidence in a group without attribute/exposure}} = \frac{a / (a+b)}{c / (c+d)}$$

		DISEASE status		Total
		yes	no	
EXPOSURE status	yes	a	b	a+b
	no	c	d	c+d
Total		a+c	b+d	a+b+c+d

53

Risk difference

- the absolute difference between two risks (or rates)

$RD = r_1 - r_0$ $\quad [a / (a+b)] - [c / (c+d)]$

		DISEASE status		Total
		yes	no	
EXPOSURE status	yes	a	b	a+b
	no	c	d	c+d
Total		a+c	b+d	a+b+c+d

54

- We can also have different strata of exposure
- We may calculate ratio measures for each strata = we compare measure of frequency in each level with measure of frequency in the baseline (unexposed) level

Example

Death rates from CHD in smokers and non-smokers by age

55

Age	Smokers rate	Non-smokers rate	Rate ratio
35-44	0.61	0.11	5.5
45-54	2.40	1.12	2.1
55-64	7.20	4.90	1.5
65-74	14.69	10.83	1.4
75-84	19.18	21.20	0.9
85+	35.93	32.66	1.1
ALL AGES	4.29	3.30	1.3

What can you say about this table?

56

Age	Smokers rate	Non-smokers rate	Rate ratio
35-44	0.61	0.11	5.5
45-54	2.40	1.12	2.1
55-64	7.20	4.90	1.5
65-74	14.69	10.83	1.4
75-84	19.18	21.20	0.9
85+	35.93	32.66	1.1
ALL AGES	4.29	3.30	1.3

The rate ratio decreases with increasing age.

***It may suggest* that the effect of smoking on the rate of CHD is higher in younger ages.**

57

Odds of disease

- We can calculate risks, risk ratio, risk difference however the analysis is often based on **ODDS RATIOS**

58

Odds of disease/survival

- related measure of disease occurrence
- for a defined population and time period

$$\text{Odds} = \frac{\text{Cases}}{\text{Non cases}}$$

=by the time of observation

59

- In many situations, it may be easier to calculate odds ratio (**OR**) which is defined as

$$\frac{\text{odds of disease among exposed}}{\text{odds of disease among unexposed}} = \frac{a/b}{c/d} \quad (\text{odds}_1)\ (\text{odds}_0)$$

$$\text{OR} = \text{odds}_1 / \text{odds}_0$$

60

		DISEASE status		Total
		yes	no	
EXPOSURE status	yes	a	b	a+b
	no	c	d	c+d
Total		a+c	b+d	a+b+c+d

$$OR = \frac{a/b}{c/d} = \frac{a \times d}{b \times c}$$

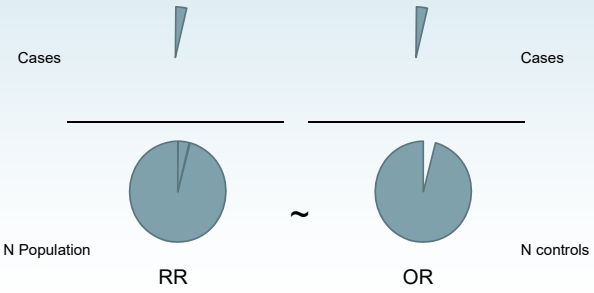
61

Odds ratio as an approximation to the risk ratio

- For a rare disease, odds ratio is approximately equal to the risk ratio (because denumerators are very similar)

62

Rare disease → OR~RR



63

If disease common:

Disease	Exposed	Unexposed	Total
Yes	50	25	75
No	50	75	125
Total	100	100	200

$\frac{a/(a+b)}{c/(c+d)}$ R_1=50/100=0.5 R_0=25/100=0.25 RR=2.0

$\frac{a/b}{c/d}$ Od_1=50/50=1.0 Od_0=25/75=0.33 OR=3.0

64

Measures of population impact

- **Population attributable risk (PAR)** is the absolute difference between the risk (or rate) in the whole population and the risk or rate in the unexposed group

$$PAR = r - r_0$$

65

Population attributable risk fraction (PAF)

- It is a measure of the proportion of all cases in the study population (exposed and unexposed) that may be attributed to the exposure, on the assumption of a causal association
- Also called the aetiologic fraction
  the percentage population attributable risk
  the attributable fraction

66

- If r is rate in the total population

$$PAR = r - r_0$$
$$PAF = PAR/r$$
$$(PAF = (r-r_0)/r)$$

67

Example

- 50 persons attended a garden party
- 25 of them developed diarrhoea in the next 3 days
- What was the risk of diarrhoea among the participants of the party?

68

- 25/50

69

Example II.

- 30 party visitors had a BBQ (minced meat)
- 24 of them developed diarrhoea
- 20 people did not eat BBQ
- 1 of them developed diarrhoea
- How would you calculate RR related to eating BBQ?

70

30 party visitors had a BBQ (minced meat)
24 of them developed diarrhoea

20 people did not eat BBQ
1 of them developed diarrhoea

- Risk among unexposed R_0:
- 1/20
- Risk among exposed R_1:
- 24/30
- Relative risk $RR=R_1/R_0=(24/30)/(1/20)=16$

71

Introduction to epidemiological study design

72

What is the purpose of Epidemiology?

"the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the prevention and control of health problems"

= Last's Dictionary of Epidemiology

It enables us to:

- Describe patterns of disease in populations
- Study the determinants or risk factors of disease
- Compare disease between groups
- Assess the effectiveness of interventions

73

Basic tool in epidemiology is the study

- foundations of a good research proposal is the **study design**
- defines how data or evidence is collected and can be used to compare disease between groups

74

Study

*"An **epidemiological study** is a statistical study on human populations, which attempts to link human health effects to a specified cause"* (wikipedia.org).

• Epidemiology studies *populations*, not individuals

• *Statistical study*: requires large number of people

• *Effects*: often means associations but here it means consequences
(i.e. disease, health condition)

• *Cause:* often means *risk factor*, because *cause* implies *causal association* which is very difficult to demonstrate in epidemiology

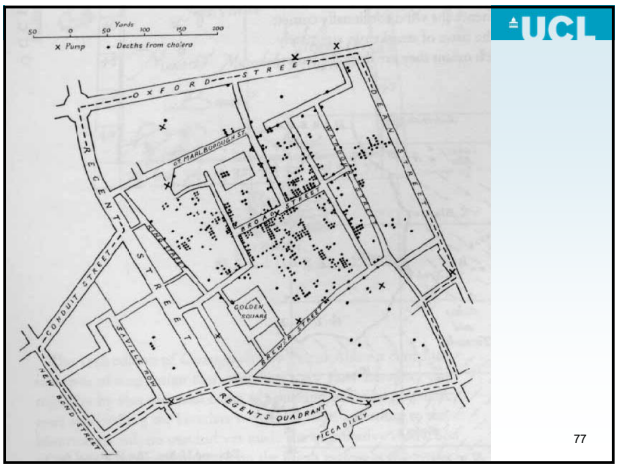
Exercise
John Snow and cholera

- **Introduction**

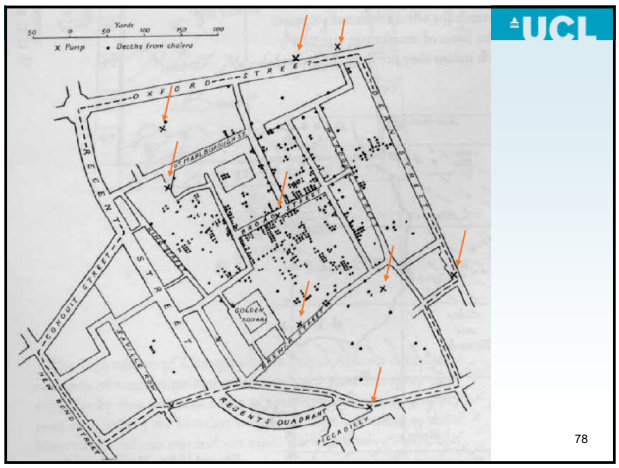
John Snow (1813-1858) was a physician in London who was distinguished for, among other things, administering chloroform to Queen Victoria at the birth of two of her children. He is best known for his studies of cholera, in particular of two outbreaks which occurred in London in 1848-49 and 1853-54.

- **Background: the 1848/49 cholera epidemic in London**
- Cholera periodically swept across Europe during the nineteenth century
- After a severe epidemic in 1832, the disease next appeared in London in 1848.
- The severity of this epidemic (approximately **15,000** recorded deaths from cholera) led to considerable discussion in the medical press. Mortality was particularly severe in the **low-lying areas along the banks of the Thames River**; hypotheses about what caused cholera included living in lower regions and the existence (contested at the time!) of microbes.

76



77



78

Classic exercise about a classic incident:

London in 1850:

- *no electricity*
- *no tarred roads*
- *full of horses and cows*
- *low standard of hygiene*
- *squalor*
- *sewage drained into the Thames*

Q1. Relevance of data to Snow's hypothesis:

Note the thoroughness of the study - data on 330/334 = 99% deaths.
A huge effort.
Appear consistent with Snow's hypothesis but need rates.
It is possible that the ratio of the number of houses or persons supplied with water by Southwark +Vauxhall (S+V) compared to number supplied by Lambeth (L) is 286/14, thus the numbers of deaths are as expected.

Q2. Deaths per houses data

- He inferred that these data supported his hypothesis that cholera was transmitted.

Risk S+V = 1,263/40,046 = 31.54 per 1000 houses

L = 98/26,107 = 3.75 per 1000 houses

Other = 1,422/256,423 = 5.55 per 1000 houses

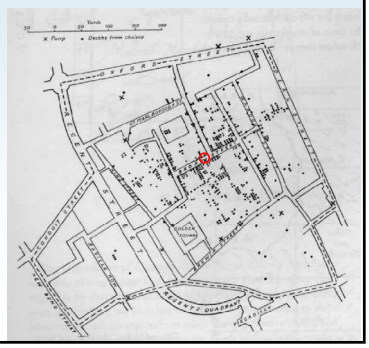
- <u>But we should also consider:</u>

a) perhaps S+V houses were bigger, divided into flats?

b) perhaps S+V houses in poorer, lower, denser area? Note that S+V area is lower lying, along the river.

Better data because he obtained denominators. Thus, he could look at deaths in relation to the number of houses receiving water from each company. Although he did not know the number of *people* actually living (and drinking) in each house, it was the best approximation he could get of the number of people "at risk".

Q3. The spatial clustering of cases suggests that the pump in Broad Street is the source – the density of cases decreases in all directions from this pump. However, note that (again) no denominators are given – how many people lived near to the Broad Street pump compared to the other pumps?



Q4. Pump closed on the 8th - epidemic was almost over by then.

So, it was not the removal of the pump handle that caused the epidemic to stop (although this had great publicity value).

BUT Epidemic may have stopped as a result of:

- exhaustion of susceptible (local people had already become ill or had fled)

- dilution of contamination

Epidemiology = comparison

- 550 cases of stomach cancer

- 550 cases of stomach cancer in Hertfordshire

85

- 550 cases of stomach cancer in Hertfordshire in 2005

86

- 550 cases of stomach cancer in Hertfordshire in 2005
- Population 550,000

=> Rate

87

- 550 cases of stomach cancer in Hertfordshire in 2005
- Population 550,000

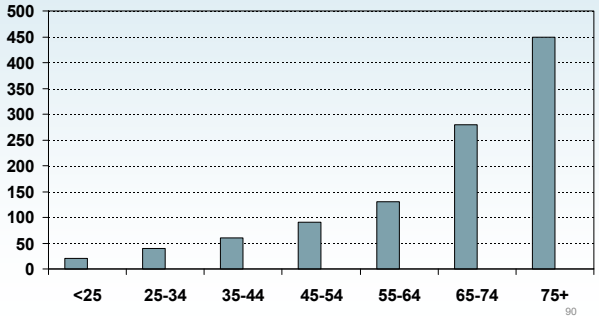
=> Rate 100/100,000

88

What else would be our interest?

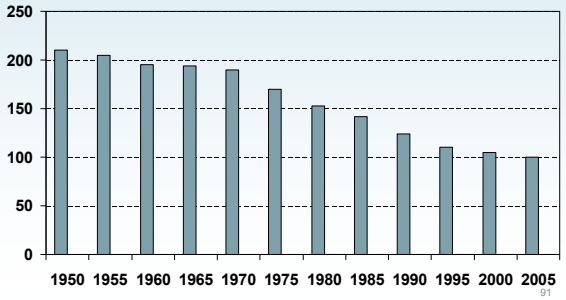
89

Stomach cancer by age group, 2005, per 100,000



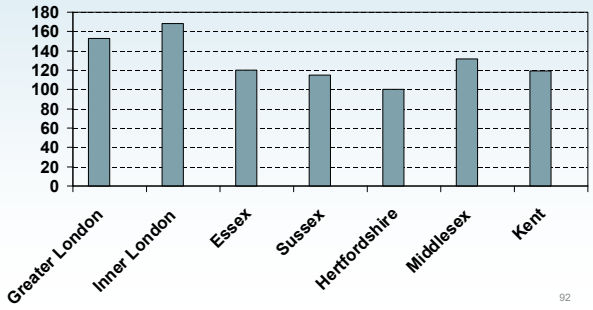
90

Stomach cancer in Hertfordshire, 1950-2005, per 100,000

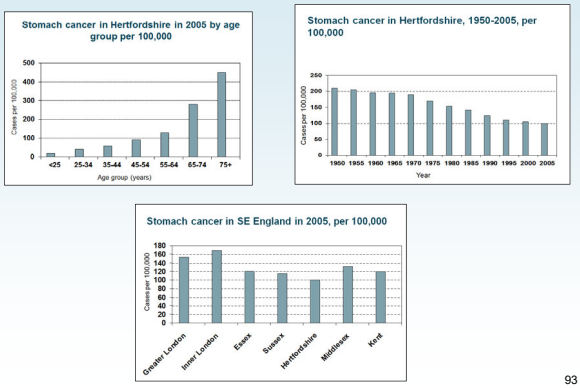


91

Stomach cancer in SE England in 2005, per 100,000



92



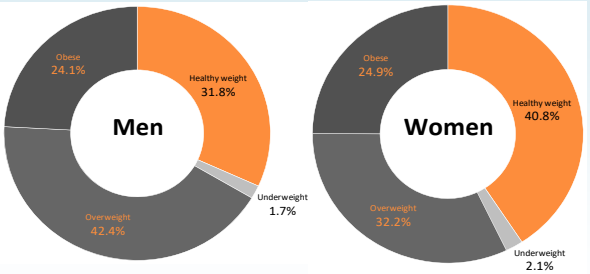
93

We compare

94

Adult prevalence by BMI status

Health Survey for England (2008-2010 average)

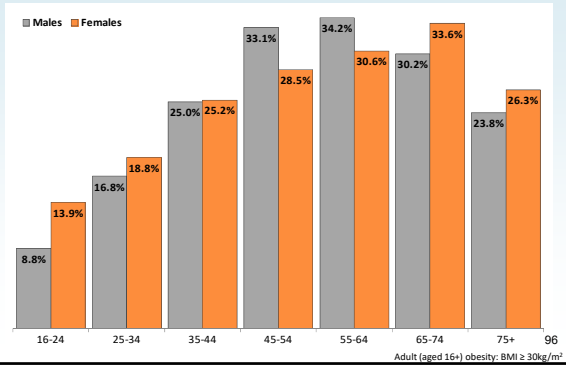


Adult (aged 16+) BMI thresholds
Underweight: <18.5kg/m²
Healthy weight: 18.5 to <25kg/m²
Overweight: 25 to <30kg/m²
Obese: ≥30kg/m²

95

Adult obesity prevalence by age and sex

Health Survey for England 2008-2010

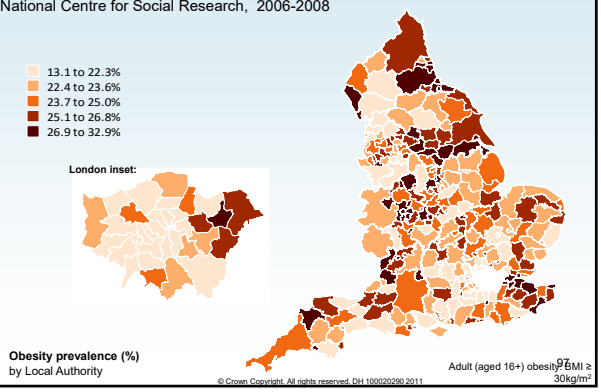


96

Adult (aged 16+) obesity: BMI ≥ 30kg/m²

Adult obesity prevalence modelled estimates

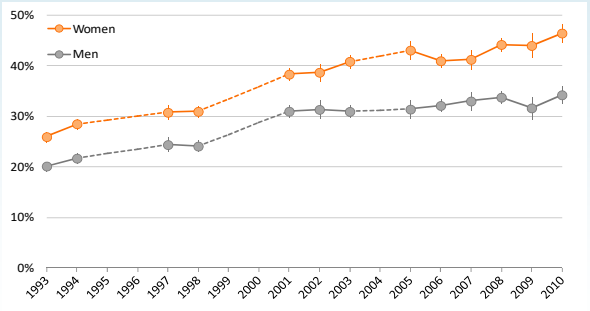
National Centre for Social Research, 2006-2008



97

Trend in raised waist circumference among adults

Health Survey for England, 1993 - 2010



The chart shows 95% confidence limits

Adults aged 16+ years

Raised waist circumference defined as >102cm for men and >88cm for women

98

Epidemiology = comparison

- Type of comparison (= type of study) depends on purpose.
- E.g.
  - ***Describe*** the disease / condition
  - Study (***analyse***) its determinants / causes
  - Study (***analyse***) prevention / treatment

99

Two primary criteria

- Descriptive vs. analytical
- Observational vs. interventional

100

Descriptive vs. analytical studies

- describe a pattern of occurrence of a disease: ***descriptive studies*** (always observational)
- to analyse the relationship between a disease and an exposure of interest: ***analytical studies*** (can be both observational and interventional)

101

Descriptive studies

- Describe patterns of disease occurrence
- **Fast and cheap BUT often do not allow proper comparisons**
- Useful for:
  - health services planning
  - hypothesis formulation in research
- Usually based on existing data:
  - mortality
  - reporting of diseases (infections, STDs, cancers...)
  - hospital and medical records
  - Census
  - employment statistics etc.

102

Four basic questions :
WHAT ? Who? Where? When?

Person (Who?)
age, sex, marital status, social class etc.

Place (Where?)
Geography within countries (cancer atlases etc.)
or internationally
(Japanese more stomach ca than in USA)
! *Special case - migrant studies*

Time (When?)
Changes over time:

- sudden onset of diseases (thalidomide, toxic shock sy)
- seasonal pattern (births, deaths, infections, etc.)
- secular trends

All in relation to the "**What**"

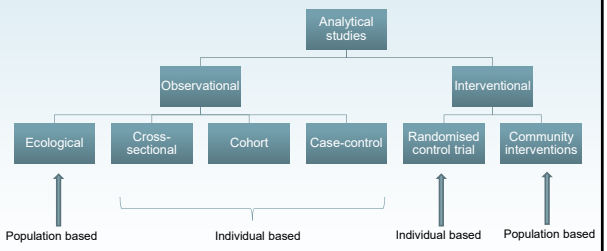
103

Analytical studies

- Analysed relationship between exposure and disease
- Often used in aetiological research
- Include
  - ecological studies
  - cross-sectional studies
  - cohort studies
  - case-control studies
  - interventional studies (RCT, prevention trials etc)

104

Analytical studies



105

Observational Studies

- Sampling
  - **Determined by** outcome and/or exposure
  - Examples of exposure: smoke, physically active, SES
  - Examples of outcome: disease or state of ill health
- Timing
  - Single point in time
  - Retrospective (CAVE how questions worded)
  - Prospective (from now → future)

106

Observational vs. interventional studies

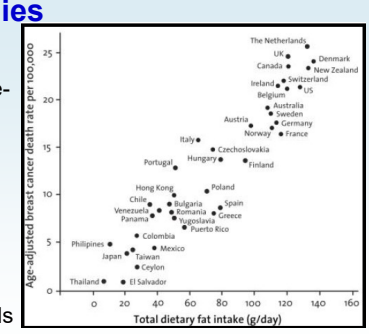
- ***Observational studies*** -observe the populations or individuals under study
  - descriptive studies
  - ecological studies
  - cross-sectional studies
  - cohort studies
  - case-control studies
- ***Interventional studies*** -where the investigators intervene, e.g. they assign exposure or a health measure to a particular individuals or groups
  - prevention studies
  - randomised clinical trials
  - community interventions

107

Ecological studies

- Grouped data
- Geographical or time-series
- Cheap & quick
- Useful to generate hypotheses

- Ecological fallacy
  = it is wrong to extrapolate from groups to individuals



Source: http://www.europeanpublichealth.com/research-methods/quantitative-or-qualitative-research/quantitative-research/

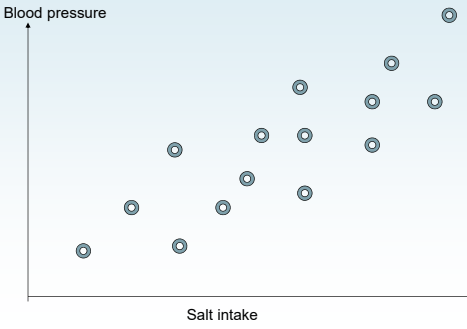
108

Ecological fallacy

- This is a logical fallacy in the **interpretation** of statistical data where inferences about the nature of individuals are deduced from inference for the group to which those individuals belong
- **Extrapolation from groups to individuals** is conceptually inappropriate
- Situation when individual-level and group-level (ecological) associations differ
- Individual data are necessary to estimate the association at the level of the individual

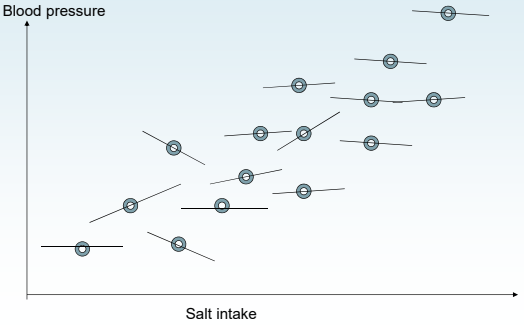
109

Ecological fallacy (1)



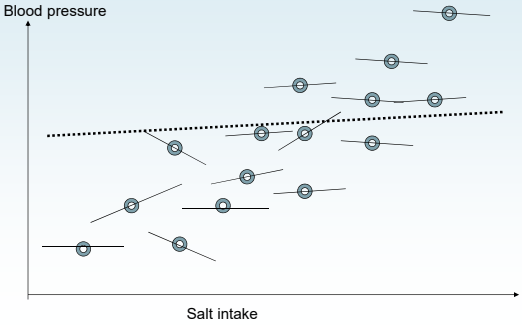
110

Ecological fallacy (3)



111

Ecological fallacy (4)



112

Example: The INTERSALT study

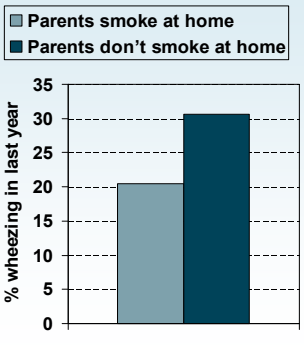
- Ecological analysis
  - Increase in salt intake by 100 mmol/day was associated with increase in SBP by 7.1 mm Hg
- Individual level analysis
  - increase by 1.6 mm Hg of SBP

From Elliott et al, BMJ 1996

113

Cross-sectional studies

- All data collected at one point in time
- Prevalence
- Relatively cheap & quick
- Useful to estimate burden of disease
- Difficult to make causal inference



114

Cross-sectional study

Survey – all measurements

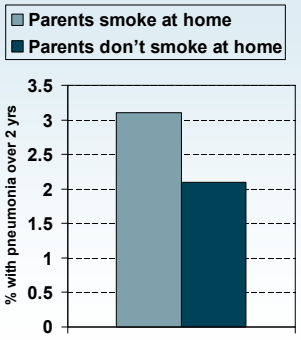
The only way to **measure** "exposures" and "outcomes" is
- **at the time of survey or**
- **retrospectively**

Time

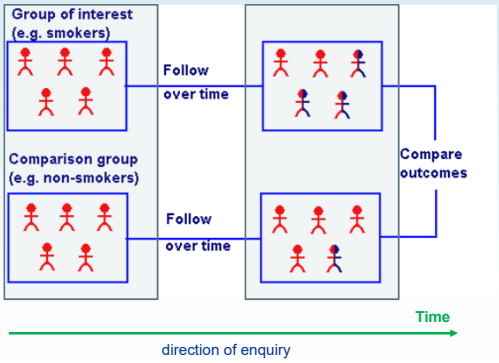
115

Cohort studies

- Exposure measured in healthy individuals
- Follow up
- Incidence
- Time consuming & expensive
- Temporality clear
- Possibly the "best" observational design



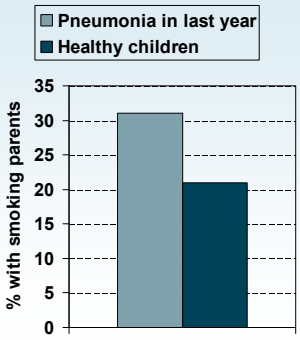
116



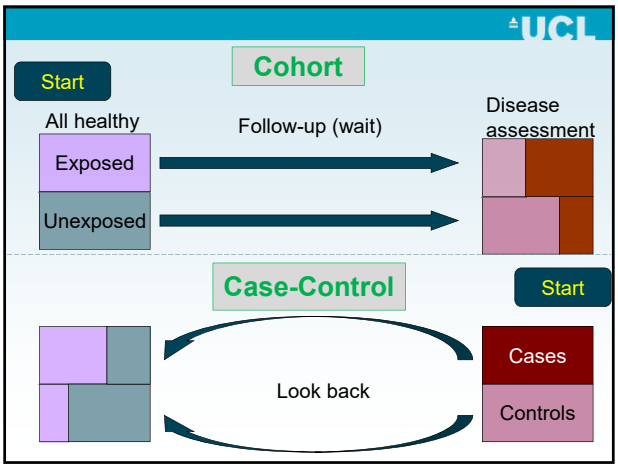
117

Case-control studies

- Cases vs. controls (current status)
- No follow up
- Asking about exposure in past
- No incidence or prevalence
- No need to wait for cases → quick
- Temporality may be a problem
- Good for exposures stable over time

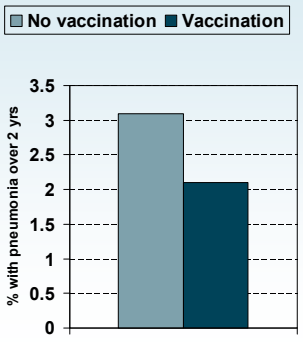


118

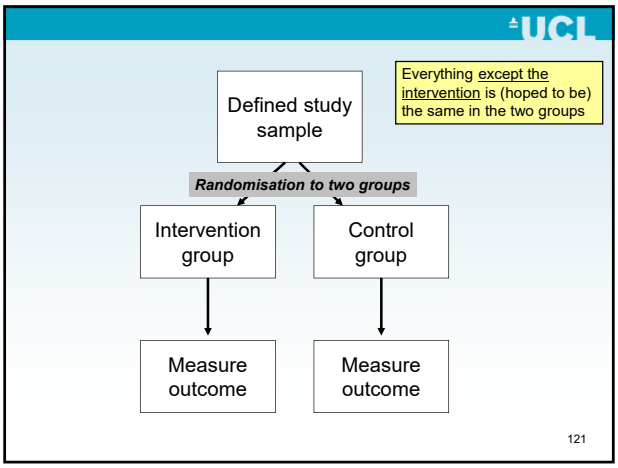


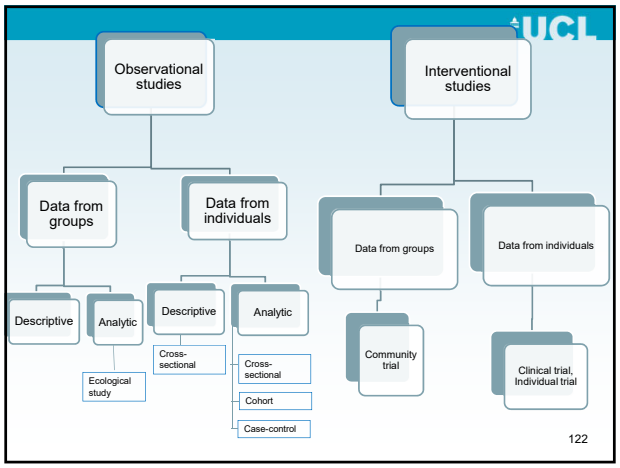
Interventional studies

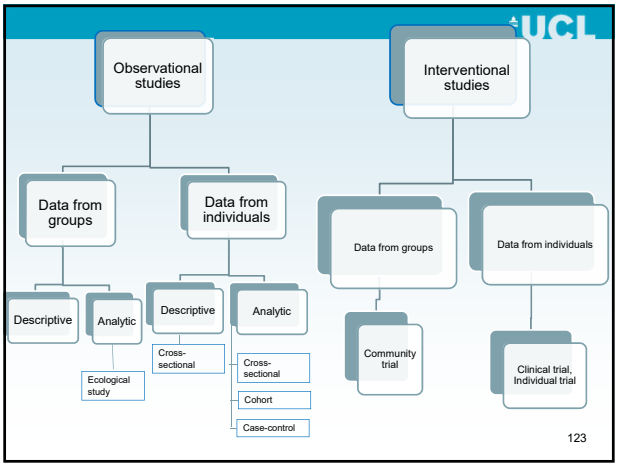
- Exposure allocated by researchers (often randomly)
- Follow up
- Incidence
- Time consuming & expensive
- Temporality clear, no confounding / bias
- Gold standard



120







Types of comparisons in different types of studies

Study design	Type of comparison
Ecological studies	Comparing disease frequency between populations
Cross-sectional studies	Comparing disease frequency between persons with and without characteristic of interest IN ONE TIME
Cohort studies	Comparing disease incidence between exposed and unexposed persons IN MORE TIME POINTS
Case-control studies	Comparing frequency of (PAST) exposure between cases and healthy controls
Interventional studies	Comparing incidence of events in persons exposed to the intervention of interest and in control group

124

Applications of different observational and analytical study designs

	Ecological	Cross sectional	Case control	Cohort
Investigation of rare disease	++++	-	+++++	-
Investigation of rare exposures	++	-	-	+++++
Examining multiple outcomes	+	++	-	+++++
Studying multiple exposures	++	++	++++	+++
Measurement of time relationships between expo and outcome	+	-	+	+++++
Direct measurement of incidence	-	-	+ [1]	+++++
Investigation of long latent period	-	-	+++	+++ [2]

1 incidence only if the sampling fraction known for both cases and controls
2 if historical cohort

125

Summarization I.

126

What types of studies do you know

127

Exercise

Study design overview
10 mins

Exposure= independent variable (e.g., smoking)

Outcome= dependent (e.g., lung cancer)

128

Rates

- What can you name and define?

129

What types of studies do you know

130

What does it means when we say ‘retrospective’ and ‘prospective’?

- What study uses which approach?

131

Exercise

Study design overview
10 mins

Exposure= independent variable (e.g., smoking)

Outcome= dependent (e.g., lung cancer)

132

a) 100 cases of stomach cancer reported to local cancer register were compared to 200 healthy subjects randomly selected from population register. 60 patients with cancer and 60 healthy subjects reported frequent consumption of spicy foods.

133

•What type of study is this?

Case-control study. The clue is that there were 100 cases of stomach cancer and 200 healthy people.

•What was the outcome measure of disease in this study?

Stomach cancer as defined by the local cancer register

•What was the measure of exposure in this study?

Reported consumption of spicy foods, probably in a questionnaire. It should be noted that this is a subjective measure because it is self-reported.

•What measure of the association between exposure and disease can be calculated in this study?

Odds ratio

	Disease Yes (cases)	Disease No (controls)
Exposure Yes	60 (a)	60 (b)
Exposure No	40 (c)	140 (d)
Total	100	200

OR = odds of being case among exposed / odds of being case among controls = (60/60)/(40/140) = 3.5

OR = (a/b) / (c/d) = (a/b) x (d/c) = (a x d) / (b x c) = (60x140) / (40x60) = 3.5

Interpretation: there are 3.5 times the odds of stomach cancer among those frequently eating spicy food compared to those that do not eat spicy food

134

b) 160 men HIV positive were divided into two groups (80 patients in each). One group were given a new drug, the second were given a drug in common use. After 1 year of this treatment, 20 patients with the new drug developed AIDS, as did 15 patients in the other group.

135

•What type of study is this?

Interventional study, possibly randomised controlled trial

•What was the measure of disease in this study?

Incidence of AIDS i.e. new cases of AIDS.

•What was the measure of exposure in this study?

A new drug

•What measures of the association between exposure and disease could be calculated in this study?

Risk ratio (RR), or absolute risk difference (ARD), or number needed to treat (NNT) or harm (NNH)

	Disease Yes (cases)	Disease No (controls)	Total
Exposure Yes	20	60	80
Exposure No	15	65	80
Total	35	125	160

*RR = (20/80) / (15/80) = **1.33** (new treatment more harmful than the old one)*

*ARD = (20/80) – (15/80) = 0.25-0.19 = **0.06** = 6% . Since this is an increased risk, we are looking at NNH not NNT*

*NNH= 1/0.06= **17.** In other words, it is predicted that there will be 1 new case of AIDS for every 17 individuals given the new drug.*

136

c) 650 subjects underwent ultrasound examination for gall stone disease; at the same time, they completed a questionnaire and anthropometric measurement (weight, height, waist and hip circumference). Obese subjects, those with BMI over 29, had 3.4 times more often gall stone disease than those with BMI less than 25.

137

- **What type of study is this?**

Cross-sectional study

- **What was the measure of disease in this study?**

Prevalence (presence of gall stones at ultrasound examination) in the 650 subjects selected.

But unclear if 650 subjects were randomly selected from general population

- **What was the measure of exposure in this study?**

Measure of exposure = being obese defined as a BMI over 29.

- **What was the measure of the association between exposure and disease in this study?**

Prevalence ratio, comparing prevalence of gall stones among obese and non-obese subjects.

*NOTE: BMI=weight/(height*height)*

Usual ranges: 20-24.9= normal, 25-29 = overweight, 30+ = obese

138

d) 17,000 men and women underwent an examination and completed a questionnaire. On the basis of the questionnaire, the subjects' jobs were classified as high stress (5000 subjects) and low stress (12000). Over the next 7 years, 50 persons in high stress jobs and 120 persons in low stress jobs developed a heart attack.

•What type of study is this?
Cohort study
•What was the measure of disease in this study?
Incidence of heart attack
•What was the measure of exposure in this study?
Self-reported stress at baseline, classified as high or low stress
•What measure of the association between exposure and disease could be calculated in this study?
Risk ratio

	Disease Yes	Disease No	Total
Exposure Yes	50		5000
Exposure No	120		12000
Total			17000

RR = [50/5000] / [120/12000] = 0.01 / 0.01 = 1 (no effect)

Summarization II.

Rates

- What can you name and define?

Exercise

-Risk exercise

-Ignaz Semmelweis

15 mins

Exercise

1,000 of retired police workers was followed for 25 years. Half of them were regular alcohol drinkers, and there were 20 cases of liver cancer in this group. In the rest of the group, there was found 10 cases of the cancer.

i) Build up the table and calculate absolute risk for each group
ii) What is the relative risk/risk ratio among regular drinkers in comparison with others?
iii) Calculate odds ratio for the same association. What do you think about results ii) and iii) ?

i)

	Cancer	No cancer	Total	Rel. risk
Regular drinkers	20	480	500	20/500=0.04
Non regular/non drinkers	10	490	500	10/500=0.02
Total	30	970	1000	

ii) **What is the relative risk among regular drinkers in comparison with others?**

RR=0.04 / 0.02= 2.00

Dates of study entry, diagnosis and end of follow up (dropout or death) would be needed to calculate person-years for the denominator.

iii) **Calculate odds ratio for the same association. What do you think about results ii) and iii) ?**

***OR = a x d / b x c**= 20x490/10x480=9800/4800 = 2.04*

Results in b) and c) are very similar. We have very rare outcome in this calculation and therefore OR and RR are similar and we can say that OR is good approximation of RR

145

Puerperal fever

Ignaz Semmelweis (1818-1865) began his medical career in 1844 in obstetrics and midwifery at the Vienna General Hospital (Allgemeines Krankenhaus). There were two obstetric divisions in the hospital: patients in the first division were examined by doctors and medical students, while midwives attended to the patients in the second division. Semmelweis noticed that there were more maternal deaths in the first division than the second division.

In this exercise you will follow Semmelweis' steps investigating the problem.

146

a. Calculate the total and year specific mortality rate for the 6-year period (1841-6) in the first and second divisions (fill the empty cells in the table above).

Year	First division			Second division		
	Births	Deaths	Mortality rate	Births	Deaths	Mortality rate
1841	3036	237	0.08	2442	86	0.04
1842	3287	518	0.16	2659	202	0.08
1843	3060	274	0.09	2739	169	0.06
1844	3157	260	0.08	2956	68	0.02
1845	3492	241	0.07	3241	66	0.02
1846	4010	459	0.11	3754	105	0.03
TOTAL	20042	1989	0.1	17791	696	0.04

b. Do you agree with Semmelweis' claim that there were more deaths in the first division?

c. Is it necessary to calculate the mortality rates for each year in order to compare the two divisions?

147

Year	Births	Deaths	Mortality rate
Jan-April 1846	1193	194	0.16
May-Aug 1846	1039	140	0.13
Sep- Dec 1846	1120	125	0.11
Jan-Apr 1847	1240	84	0.07
TOTAL	4592	543	0.118
INTERVENTION			
May-Aug 1847	1076	50	0.05
Sep-Dec 1847	1059	42	0.04
Jan-Apr 1848	1155	14	0.01
May-Aug 1848	1107	7	0.006
TOTAL	4397	113	0.0257

e) Was Semmelweis' intervention successful?

f) Briefly comment on the importance and implications of this finding in terms of epidemiology and clinical practice.

148

We have study, we have basic results from analysis…

.....we must know how to interpret findings

149

Three major issues in interpretation of any epidemiological study

- Chance (random variation) – statistics
- Bias (i.e. systematic error)
- Confounding

- Only if all of these have been excluded, you may start thinking of a causal association

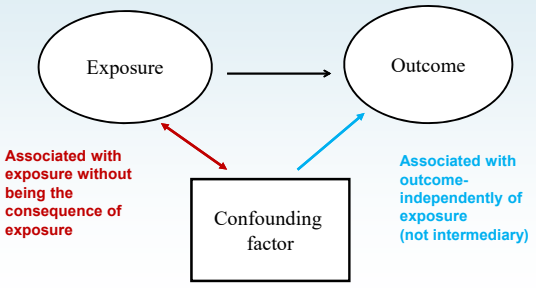
150

Confounding

- Situation when a third factor is associated with both exposure and disease
- Association between "exposure" and disease may not be causal; instead, it is due to a third factor which is associated with both exposure and disease.

151

Confounding



152

EXAMPLE
Case-control study of alcohol and lung cancer

	Alcohol	No alcohol
Cancer	450	300
No cancer	200	250

$$OR = \frac{a/b}{c/d} = \frac{a \times d}{b \times c}$$

Estimated odds ratio =1.875

153

The same data stratified by smoking:

	Non-smokers		Smokers	
	Alcohol	No alcohol	Alcohol	No alcohol
Cancer	50	100	400	200
No cancer	100	200	100	50
Estimated odds ratio	1.0		1.0	

154

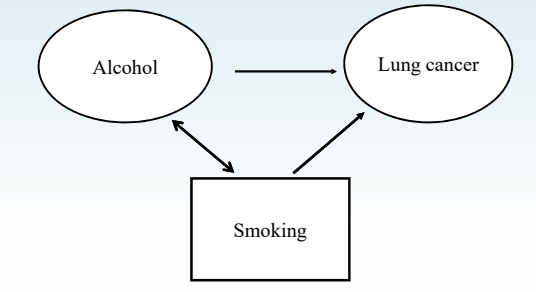
Alcohol and smoking in controls (=no cancer)

	Alcohol	No alcohol	Total
Smokers	100	50	150
Non-smokers	100	200	300
Total	200	250	450

Non-drinkers: 1 in 5 were smokers…..50 from 250
Drinkers: 1 in 2 were smokers…..100 from 200

155

Confounding



156

Most common confounders:

- Gender (men have higher mortality and more risk factors; women higher morbidity)
- Age (risk of most diseases increases with age)
- Socioeconomic status (risk of most diseases higher in lower SE groups)
- Ethnic group
- Smoking
- Alcohol
- etc...

157

Control of confounding

Design
- Randomisation
- Restriction
- Matching

Analysis (if data collected)
- Stratification
- Regression modelling

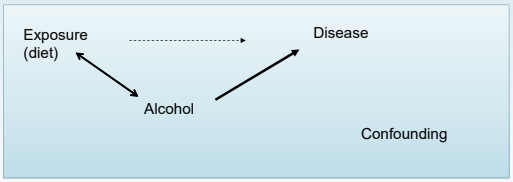
158

Residual confounding

- Unmeasured confounding factors or measurement error in confounding factors may lead to residual confounding.
- The possibility of residual confounding cannot be completely eliminated in observational studies

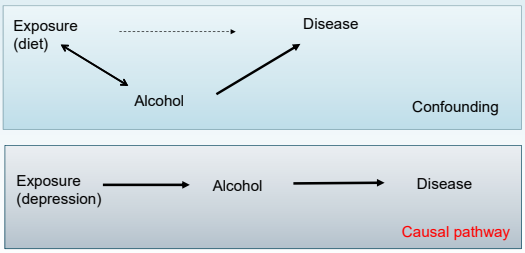
159

Confounding vs. causal pathway



160

Confounding vs. causal pathway



Statistically, confounding is the same as causal pathway
The difference is conceptual – i.e. in your head!

161

Effect modification (interaction)

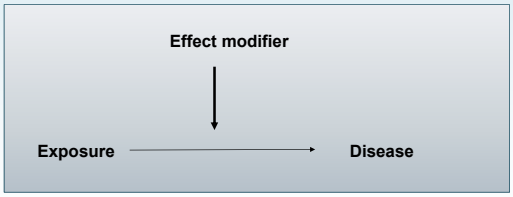
- the effect of exposure on disease is dependent on the level of a third factor

or

- a moderator specifies on whom or under what conditions another variable (exposure) will operate to produce the disease.

162

Effect modification



163

Finding out the different influence in different strata

=exploring association between exposure (independent variable) and outcome (dependent variable) within different strata of the 3rd characteristic

age groups
sex
achieved education level
geographical area

164

Example from yesterday

Death rates from CHD in smokers and non-smokers by age

165

Age	Smokers rate	Non-smokers rate	Rate ratio
35-44	0.61	0.11	5.5
45-54	2.40	1.12	2.1
55-64	7.20	4.90	1.5
65-74	14.69	10.83	1.4
75-84	19.18	21.20	0.9
85+	35.93	32.66	1.1
ALL AGES	4.29	3.30	1.3

The rate ratio decreases with increasing age.
***It may suggest* that the effect of smoking on the rate of CHD is higher in younger ages.**

166

EXAMPLE
CHD, smoking and age in British doctors study (rates per 100,000)
=Framingham study

	Non-smokers		Heavy smokers	
	Rate	**RR**	**Rate**	**RR**
<45	7	1.0	104	14.9
45-54	118	1.0	393	3.3
55-64	531	1.0	1025	1.9

167

Positive and negative effect modification

- Positive:
  - "susceptibility factor" or "vulnerability factor",
  - its presence (or higher values) strengthens the association between exposure and disease.
- Negative:
  - "resiliency factor" or "buffering factor"
  - its presence (or higher values) weakens the association between exposure and disease

168

Identification of effect modification

- Stratified analysis
- Compare effect estimates in strata
- Assess differences in effects by significance tests (p-value for heterogeneity)
- Pooled estimates (e.g. standardised) **not appropriate** when there is an interaction
- *Please note that genuine & meaningful interactions are rare*

169

Confounding vs. interaction

Confounding	Effect modification
• Alternative explanation • Distorts the "truth" • Efforts to remove it to get nearer to the "truth" • When present, stratum specific effects are similar to each other but different from the overall crude effect.	• One factor modifies effect of another factor • It is genuine, not artefact • Property of the relationship between factors • We should detect and describe it but not remove it.

170

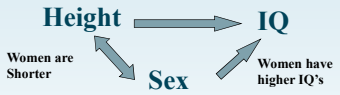
Example: Height and IQ – real association or not?

Height → IQ

Found: High negative association between height and IQ

171

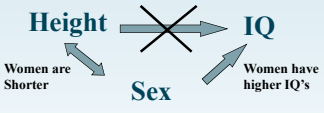
Height and IQ



- Find out that Sex is related to Height and that Sex is related to IQ
- Therefore, Sex is a ***potential*** confounder

172

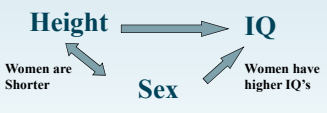
Height and IQ



If after adjustment for Sex there is NO association between height and IQ, then Sex was a confounder

173

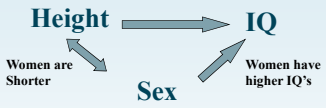
Height and IQ



If after adjustment for Sex there is still a strong negative association between Height and IQ, then Sex is not a confounder

174

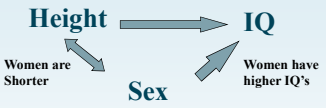
Height and IQ



If, after adjustment for Sex, there is still an association between Height and IQ, but the nature and/or strength of the association changes with Sex (=different for M and F), then Sex is an Effect Modifier.

175

Height and IQ



If there is no association between Sex and IQ, then
Sex cannot be a confounder
Likewise, if there is no association between Sex and height, then Sex cannot be a confounder

The confounder must be related to both Exposure and Outcome

176

Three main categories of alternative explanation

- Chance - random error
- **Bias** - systematic error
- Confounding – third factor explaining an association

177

Validity

- A study's results and conclusions are **valid** when they reflect the true relationship in the study population
- To assess the validity of findings we need to consider **alternative explanations** for the observed associations

178

Bias

- is a systematic error in the design of an epidemiological study which leads to a distortion or error in the study results
- an association will allow to be distorted if error is differential

179

Bias can affect

- Estimate of one variable
- Estimate of association between variables

180

Errors (biases) may be

a)

- Non–differential vs. differential
  - error in one variable not related to / dependent on the value of other variables
  - error in one variable is related to value of other variable

181

Example: sex differences in HDL-cholesterol

non-differential – badly calibrated measurement of HDL-cholesterol does not bias estimate of mean sex difference (the error cancels out)

differential - measurement of HDL-cholesterol in different single sex studies using different labs: biases estimate of mean sex difference – unless labs carefully calibrated against an external standard.

E.g. cases and controls analysed in different labs!

182

Errors (biases) may be

b)

- Selective vs. informative
  - Related to selecting subjects into study
  - Related to collecting information

183

Selection bias

- due to errors in the way sample is recruited
- a distortion that results from procedures used to select subjects or their participation
- resulting in a difference in the characteristics between those who are included in the study and those in study population but not included in the study sample

184

- The study sample
  - representative or random sample better than volunteers
  - high response rate (>70%)
- Follow-up participation in longitudinal study
- Item non-response

If non-response is related to the exposure and/or outcome, then the study may produce invalid findings

e.g. sick smokers may refuse to participate more often than healthy smokers

185

Particular concern in case-control studies because **exposure and disease are both present** at time of recruitment

Hospital-based studies are problematic because cases are filtered: not all cases go to hospital, not all cases get the correct diagnosis

e.g. a hospital-based study of depression will involve severe cases only

186

Information bias

- due to errors in way in which information collected from the sample
- errors in the way information about exposure or disease collected

=>misclassification - putting subjects in wrong category
inaccurate estimates of occurrence of effect size, or even direction of association

e.g., exposed as unexposed, case as control

187

Important types of information bias include

- **Reporting/recall bias**: by study participants
- **Observer bias**: in measurements by research personnel
- **Diagnostic bias**: probability of detection or correct identification of disease across study groups or over time

188

Aroused misclassification may be

- **Random** - above / below
- **Systematic** – all in one direction

- **Non–differential** (error in one variable not related to / dependent on the value of other variables)
- **Differential** (error in one variable is related to value of other variable

189

Non-differential misclassification:

- Tend to bias estimates towards null
- Cholesterol machine giving random readings
- Underestimated effect traditionally seen as less of problem than overestimate

190

Differential misclassification

- Can distort associations, and can produce spurious associations

191

i) Reporting bias

- May underestimate some behaviours eg alcohol, smoking
- In CS or CC studies when exposure & disease assessed at same time – bigger problem
- eg depression and poor physical health
- Often not conscious – placebo effect

192

ii) Recall bias

- Particular problem in case control studies or as part of retrospective part of longitudinal study
- Case may have better recall of exposure
- Eg., mothers of babies with congenital abnormality
- Diarrhoeal illness and food consumption

193

iii) Observer bias

- investigator classifies exposure differently in cases / control

or

- the investigator diagnoses disease differently in exposed / unexposed participants

=> the results are distorted

194

iv) Interviewer bias

- Interviewer may probe cases more closely for exposure
- May look for endpoint more carefully in those exposed to disease

=> **Study must be blinded**

195

v) Detection bias

- Differences may occur in accessing medical care
- Differences in diagnostic criteria
- These differences may be associated with exposure eg social class / country
- Hence detail paid to ascertainment and validation of endpoints

196

What can we do to prevent / reduce bias?

Selection bias

- random sampling from study population
- strategies to reduce non –response eg repeat mailings, offering different times at clinic
- proper choice of control group in case-control studies

197

Recall / reporting bias

- recall bias : try to obtain objective information on past exposures wherever possible or use proxy informants
- reporting bias – include lots of different questions so that subjects are hypothesis blind
- trials should be controlled and blinded

198

Observer bias

- investigators blind to case / exposure status wherever possible
- use standardised instruments and protocols, back translations
- ideally use centralised measurement or calibrate instrument
- periodic check on staff to check for differences in procedures

199

Detection / diagnostic bias

- aim for population - based ascertainment of cases
- follow 'Standardised diagnostic criteria'

200

Assessment of bias

- Non-responders questionnaire
- Baseline characteristics of those lost to follow can be analysed and compared to those remaining in study
- Objective validation of self-reported information
- Sensitivity analyses to estimate effect of bias

201

Bias: the silent menace

- Cannot be assessed numerically
- No software to identify bias
- If there is flaw in the design of the study increasing numbers will not get rid of it !
- Can only be assessed by careful evaluation of the design

202

Publication bias

High-impact journals prefer clear, positive results! ☹

Bias in systematic reviews

Form of selection bias arising if null studies are not published

If not included the overall estimate is biased upwards.

Minimised by searching grey literature, trial registers and conference proceedings to include null/negative results

e.g. the 'drug effectiveness cycle' (β-blocker-mortality example), selective serotonin reuptake inhibitors in treating depression

203

Publication bias

Failure to publish

- a negative or inconclusive trial result
- a small trial may be abandoned

Duplicate publication

- a large treatment effect
- need for research output

E.g. nine trials of ondansetron (antiemetic) in 23 (!) publications (Tramer et al BMJ 1997)

204

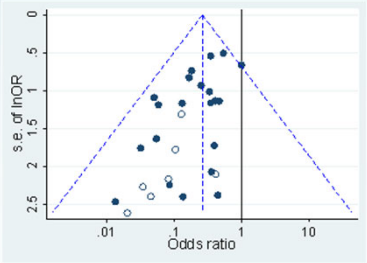
How to avoid publication bias

- To make sure studies are not double counted
- To search for unpublished studies (e.g. contact researchers directly)
- To use non-English language publications
- Statistical checking (funnel plots: smaller studies report more extreme results)
- Registration of studies and to make sure all results are in public domain (not yet fully achieved)
- Trial registration: assigns unique trial identification numbers, and to record other basic information about the trial so that essential details are made publicly available
- From 2004 International Committee of Medical Journal Editors (ICMJE) would consider trials for publication only if they had been registered before the enrolment of the first participant.

205

Funnel plot:

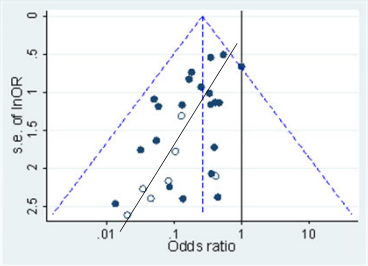
asymmetrical plot in the presence of bias: some smaller studies (open circles) are of lower methodological quality and therefore produce exaggerated effect estimates



206

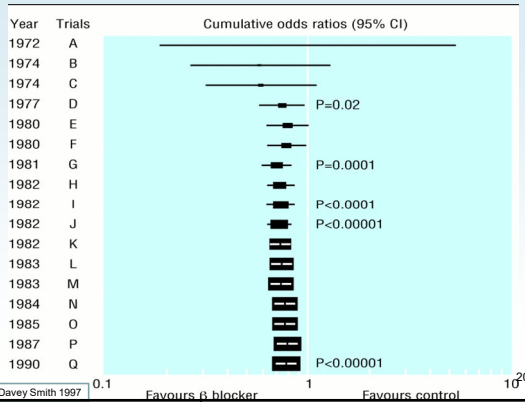
Funnel plot:

asymmetrical plot in the presence of bias: some smaller studies (open circles) are of lower methodological quality and therefore produce exaggerated effect estimates



207

Beta-blockers and total mortality after MI: meta-analysis



208

Conclusions:

- All studies are imperfect
- Most studies are subject to measurement error and various biases
- The question is: **are the results valid enough for my purpose?**

209

Three major issues in interpretation of any epidemiological study

- Chance (random variation) – statistics
- Bias (i.e. systematic error)
- Confounding

- Only if all of these have been excluded, you may start thinking of a causal association

210

Causality

1/ we find an association between exposure and outcome

2/ we need to ask whether the association is causal = does the exposure cause the outcome?

211

What is a cause?

Rothman (1986):
An event, condition, or characteristic that plays an essential role in producing an occurrence of the disease. Source - Modern Epidemiology.

- Something that has an effect
- Alters disease frequency or health status

212

Association versus Causation

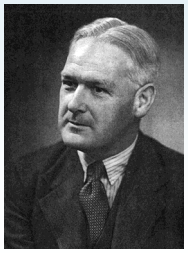
- Epidemiological research aims to discover aetiology of disease
- Epidemiology is the study of the association between a potential cause (risk factor/determinant) and a specific disease (outcome).
- Presence of a valid statistical association does not imply causality
- Association is not the same as causation
- Goes beyond association
- How do we decide whether a given association is causal or not?

213

Sir Austin BRADFORD HILL (1897-1991)

"Exposure and Disease: Association or Causation?"

1. Strength
2. Consistency
3. Specificity
4. Temporality
5. Dose-response
6. Biological plausibility
7. Coherence
8. Reversibility
9. Analogy



214

Guidelines for inferring causation

- The Bradford-Hill criteria of causation
  (J Royal Soc Med 1965; 58: 295-300)

215

Strength of association

- Measured by RR, OR
- Strong association is less likely to be due to undetected confounding or bias
- Weak association may be causal
  - Measurement error dilutes association

216

Consistency of association

- Association observed in several different studies with different study designs and populations
- Less likely that same biases present in all of them
- Inconsistency between populations may reflect lack of association or differences in the prevalence of other causal complements

217

Specificity of association

- Occurs when a single factor is associated with a single outcome
- Increasingly irrelevant to current models of disease causation (single factor many outcomes)

Example
- asbestos and mesothelioma – shown
- HIV and AIDS – shown
- Low lead exposure and IQ – not clear. IQ is not a definable brain condition so there is the potential for confounding e.g. SES

218

Temporal sequence of association

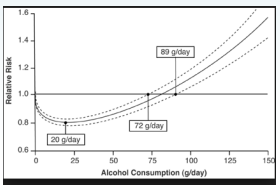
- The exposure must precede outcome
- Optimal study designs = randomised intervention study or prospective cohort study
- Weak designs for temporality: cross-sectional, case-control study
- Reverse causality may be problem in cohort or case-control study

219

Biological gradient (=dose response)

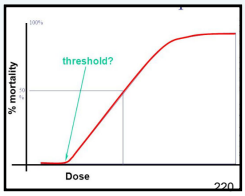
- Observe an increase in the magnitude of risk of outcome with magnitude of exposure
- Unlikely to be explained by bias or undetected confounding
- Lack of a biological gradient does not rule out causality

J- or U- shaped relationships



Source: pubs.niaaa.nih.gov

Threshold effect



220

Example to previous

Persons who have increasingly higher exposure levels have increasingly higher risks of disease

Smoking Status	Lung Cancer risk
None	1.0
Ex-smoker	1.1 (0.7-1.6)
1-20 per day	2.6 (1.7-4.0)
20-40 per day	4.4 (2.8-6.9)
40+ per day	6.8 (4.3-10.7)

221

Plausibility of association

- Practically we may accept a possible causal association even when there is no plausible mechanism or explanation
- Acceptance depends on how “unlikely association is”
- Reported association may stimulate search for mechanism

Example
- Cigarettes & lung cancer. Carcinogenic substances in cigarettes
- Low fibre diet & colon cancer. Dietary fibre increases intestinal motility and dilutes/absorbs fecal carcinogens

222

Coherence of association

- Reported association does not conflict with current knowledge
- Can lead to publication bias
- Can discourage search for alternative associations

Example
- Serum cholesterol lowering effect on heart attack, regardless of the means – diet or drug

223

Experiment (reversibility)

- Removal of exposure leads to a reduction in the risk of the outcome
- Currently perceived as the strongest type of evidence
- May be difficult to ascertain in diseases with long lag times between exposure and disease

224

Analogy

- Other similar demonstrated associations
- In practice may be limited by current knowledge

225

Bradford Hill Closing Remarks (1965)

"I do not believe … that we can usefully lay down some hard-and-fast rules of evidence that must be observed before we accept cause and effect.

None … can bring indisputable evidence for or against the cause and-effect hypothesis and none can be required …

What they can do, with greater or less strength, is to help us to make up our minds on the fundamental question - is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?"

226

Causal Inference

- Not just ticking boxes
- Weigh evidence of causal association against other explanations
- Understanding, judgement & interpretation are crucial
- Cannot prove a causal association
- Can only be inferred based on evidence
- May change in the light of new evidence

227

Reverse causality

Refers to the possibility that the link between exposure and outcome is a result of the disease or disease process being studied, not the exposure

Reverse causality is a type of confounding in the sense that it is 'real' and not an artefact of study design. It is relevant in some situations but not others

228

Example of potential reverse causality

Researchers are interested in the link between **blood levels of inflammatory markers** and **later CVD**

There are 4 possible explanations

1. Inflammation → atherosclerosis (causal association)
2. Atherosclerosis → inflammation (reverse causal association)
3. Inflammation ←→ atherosclerosis (association is bi-directional)
4. Other processes lead both to atherosclerosis and inflammation (confounding) e.g. diet

229

Public health policy

- Ideally based on 'evidence' - meta-analyses and systematic reviews
- Considerations of efficiency, cost-effectiveness and harm
- Eradication of poverty for improving health?
- Reduction in social inequality for reducing health inequality?

230

Causation and public health

- There is moment when action may be taken – it may vary from introduction of a new drug to advice to public on certain practice, or new legislation being introduced
- Complex process taking into account costs, benefits and harms
- Even when evidence become overwhelming, governments may be slow to act

231

Summary

- Epidemiology = the study of the distribution and determinants of disease in population
- Types of epidemiological studies = interventional, observational studies
- Measures of disease occurrence
- Bias, confounding, chance
- Causality

232