How do we study the causes of disease?

Seif Shaheen
Professor of Respiratory Epidemiology
Centre for Health Sciences
Barts and The London School of
Medicine and Dentistry

s.shaheen@qmul.ac.uk

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

Objectives

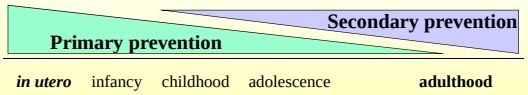
- Descriptive studies
  - Clues from geography
- Ecological studies
  - strengths and weaknesses
- Case-control studies
  - strengths and weaknesses
- Measurement of risk
  - Odds ratio
- Confounding and bias

What is epidemiology and why do it?

- The study of the distribution of disease in populations and ***factors determining the distribution***.

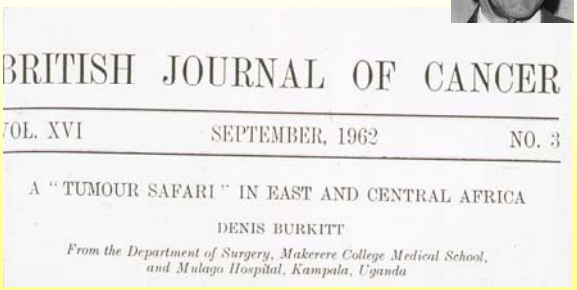
- Find causes → Prevent disease → Improve PH

Prevention



in utero infancy childhood adolescence **adulthood**

An inspiring story.....



BRITISH JOURNAL OF CANCER

VOL. XVI SEPTEMBER, 1962 NO. 3

A "TUMOUR SAFARI" IN EAST AND CENTRAL AFRICA

DENIS BURKITT

From the Department of Surgery, Makerere College Medical School, and Mulago Hospital, Kampala, Uganda

Epidemiology on a budget....

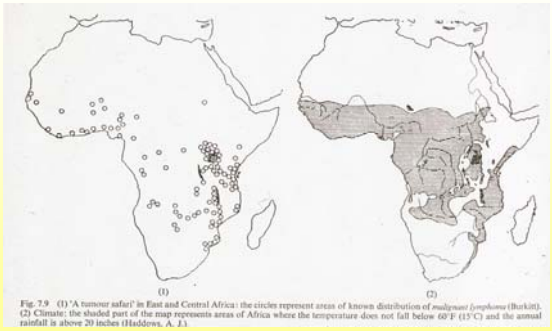
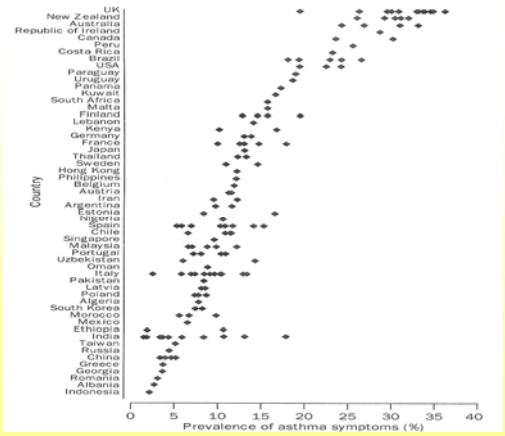
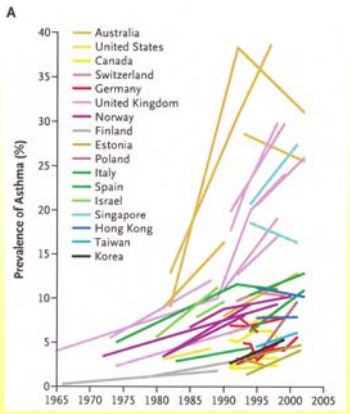
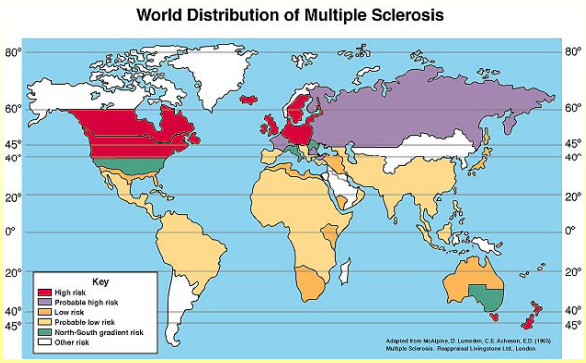


Fig. 7.9 (1) 'A tumour safari' in East and Central Africa: the circles represent areas of known distribution of *malignant lymphoma* (Burkitt). (2) Climate: the shaded part of the map represents areas of Africa where the temperature does not fall below 60°F (15°C) and the annual rainfall is above 20 inches (Haddow, A. J.).

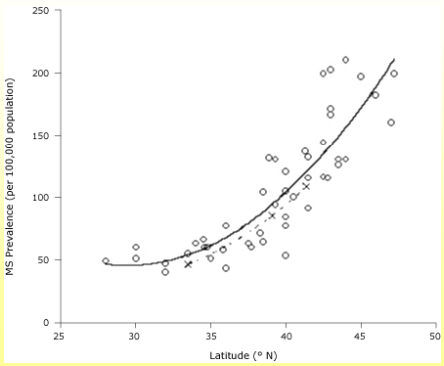




NEJM 2006; 355: 2226-35

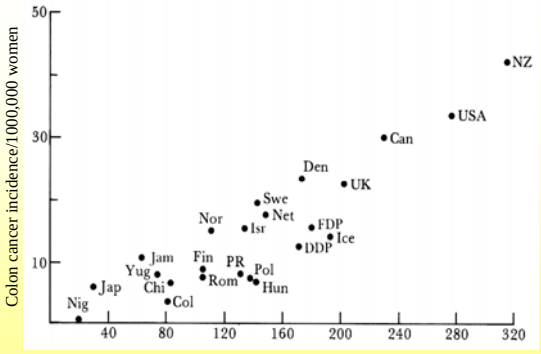


Prevalence of MS within the USA



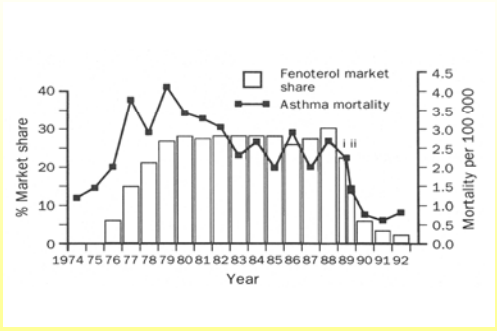
Ecological studies

- 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)



Meat : Per Capita daily consumption - grams

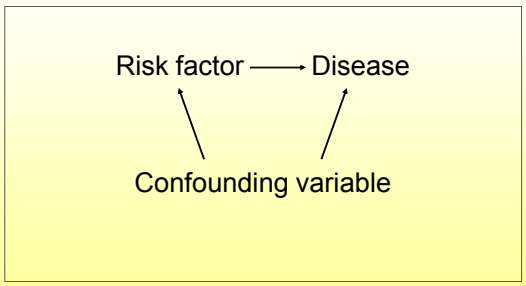
Relation between fenoterol sales and asthma mortality over time



Ecological studies

- Strengths
  - quick and cheap to do
  - generate new hypotheses / identify new risk factors
  - maximise variation in exposure
- Limitations
  - associations apply to aggregates of people but may not apply to individuals
  - difficult to allow for confounding

Confounding



Confounding: example

Meat intake ---------> Colon cancer
\ /
\ /
???

Confounding: example

Meat intake ---------> Colon cancer
\ /
\ /
? Fibre intake
? Fruit and vegetables
? Smoking

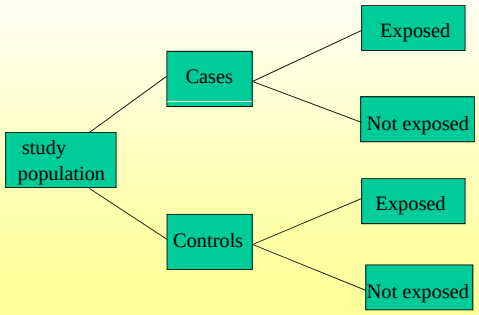
Case-control studies

- Hard to do well, easy to do badly

"…. many studies have been conducted by would-be investigators who lack even a rudimentary appreciation of epidemiological principles…….often the results are wrong because basic research principles have been violated".

Kenneth Rothman

Case control study: design



Conducting a case-control study: five steps

- Define study population (source of cases/controls)
- Define and select cases
- Define and select controls
- Measure exposure
- Estimate disease risk associated with exposure

Source of cases

- Hospital based
  - Cases from selected hospital(s) over defined period
    - Easier, cheaper; more severe disease
- Population (community) based
  - All cases (defined period/area) or random sample
    - Avoids selection factors influencing referral to hospital; less severe disease

Type and definition of cases

- Incident cases preferred to prevalent cases
  - Exposures (eg lifestyle habits) may *change* as a result of early disease
- Case definition
  - strict diagnostic criteria for presence of disease
    - Standardised / validated
  - Homogeneous
    - Nb Different phenotypes have different aetiology

Finding cases

- Ascertainment
  - Death certificates
  - Disease registers; medical records
  - Population survey
- If rare disease may have to find from large area / over many years

Sources of controls

- Hospital
  - Different diseases from cases
  - Pros
    - Same selection factors as hospital cases
    - Similar motivation/recall as cases
- General population
  - Healthy or with other diseases
  - If cases from general population
  - May use as well as hospital controls
  - Cons
    - Lower motivation/poorer recall/response rates

Defining and selecting controls

- Control definition
  - strict criteria for absence of disease of interest
- Selection of controls (*sample* of all controls)
  - must represent the population from which the cases came
    - Could have been included as cases if had developed the disease of interest
- Ratio of controls:cases
  - Usually 1:1
  - If cases limited can go up to 4:1 to increase power

Measuring exposure

- Exposure information
  - Records
  - Questionnaire
    - Recall risk factors / exposures in the past
  - Blood measurements
- Must be collected in a comparable way for cases and controls

Comparing odds of exposure in cases and controls

- **Odds of exposure**
  = number of individuals exposed / number of individuals not exposed
- **Odds ratio** = odds of exposure in cases / odds of exposure in controls

Calculating the odds ratio

		Disease outcome	
		Present	Absent
Risk factor	Present	a	b
	Absent	c	d

Odds ratio = $\frac{a/c}{b/d} = \frac{ad}{bc}$

Hepatitis B infection and hepatocellular carcinoma in the Gambia
Hepatology 2004; 39: 211-9

		HCC	
		Cases	Controls
Hepatitis B sAg	Positive	106 (56%)	62 (16%)
	Negative	82	338
	Total	188	400

Univariate Odds Ratio = $\frac{106/82}{62/338} = \frac{1.293}{0.183}$ = **7.07**

How to deal with confounding

- Matching
  - Eg match cases and controls for age, sex
  - Disadvantage: can't assess effects of these factors
- Stratification
  - Eg if effects seen in non-smokers, smoking can't confound
- **Multivariate analysis**
  - Multiple logistic regression

Leukaemia near nuclear plants

- La Hague: nuclear waste reprocessing plant
- 1978-1993: 27 cases of leukaemia < 25 yrs old
- 192 controls (up to 10 per case)
  - recruited from GP's
  - matched for sex, age, place of birth, place of residence
- Parents interviewed about risk factor exposure

BMJ 1997; 314: 101-6

Leukaemia near nuclear plants

	Leukaemia Cases	**Controls**	OR (95% CI)
Rec activity on local beaches			
< once/month	10	110	1.0
≥ once/month	17	82	**2.9 (1.1 - 8.7)**

Cellular phones and brain cancer

- 5 US hospitals
  - 1994-1998
- 469 cases of primary brain cancer
- 422 controls without brain cancer
  - hospital patients with other diseases
- Interview (questionnaire)
  - use of cellular phones

JAMA 2000; 284: 3001-7

Cellular phones and brain cancer

		Brain cancer	
		Cases	**Controls**
Cell phone use	**Yes**	66	76
	No	403	346

Odds ratio = $\frac{66/403}{76/346}$ = **0.75 (0.51 to 1.09)**

Cellular phones and brain cancer

Cell phone use (years)	**Cases** n (%)	**Controls** n (%)	OR* (95% CI)
0	403 (86)	346 (82)	1.0
1	21 (5)	30 (7)	0.7 (0.4 - 1.3)
2-3	28 (6)	24 (6)	1.1 (0.6 - 2.0)
4+	17 (4)	22 (5)	0.7 (0.4 - 1.4)

*adjusted for confounders

Selenium intake and asthma

Am J Respir Crit Care Med 2001; 164: 1823-28.

Intake/day	OR*	(95% CI)
1	1.0	
2	0.95	(0.66 to 1.36)
3	0.69	(0.46 to 1.03)
4	0.53	(0.34 to 0.81)
5	0.56	(0.35 to 0.89)

*adjusted odds ratio

p trend 0.0015

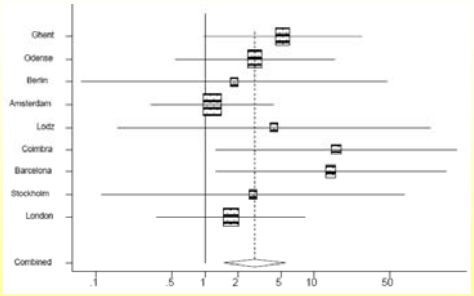
Paracetamol use and asthma

Thorax 2000; 55: 266-70.

	Cases	**Controls**	
Freq.	n (%)	n (%)	Adj OR (95% CI)
never	98 (15)	153 (17)	1.00
<monthly	259 (39)	424 (47)	1.06 (0.77 - 1.45)
monthly	172 (26)	219 (24)	1.22 (0.87 - 1.72)
weekly	105 (16)	97 (11)	1.79 (1.21 - 2.65)
daily	30 (5)	17 (2)	2.38 (1.22 - 4.64)

p trend 0.0002

Relation of paracetamol use to asthma across GA²LEN centres



Odds ratio comparing weekly versus <weekly use

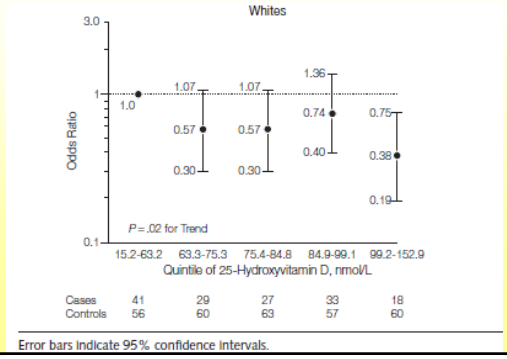
Eur Respir J 2008 ; 32: 1231 - 1236.

Nested case control studies

- "Nested" within a cohort study
- Example: prospective cohort study
  - Does low blood selenium predict ↑ risk of lung cancer?
  - Blood samples taken at baseline and frozen
  - Follow-up and collection of mortality data
  - At end of study define cases and controls
  - Measure selenium in stored samples of cases and sample of controls only
- More efficient for costly exposure measurements

Multiple sclerosis and vitamin D status in military personnel

JAMA 2006; 296: 2832-8



Case control studies

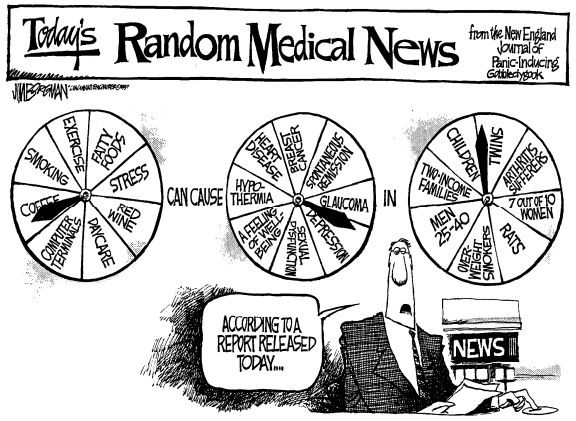
- Strengths
  - quicker and cheaper than cohort studies
  - study rare diseases
  - study multiple risk factors
  - study diseases with long latent period
- Limitations
  - prone to selection and recall bias
  - inefficient for rare exposures
  - may be difficult to establish **temporality**

Reverse causation?

Low blood antioxidants → Lung cancer ?

or

Lung cancer → Low blood antioxidants ?



Interpretation of observational study findings

- Are the statistical findings valid?
  - **Chance**?
    - What is level of statistical significance (P value)?
  - **Bias**?
  - **Confounding**?
    - Was this adequately addressed in design and analysis?
- Are the findings generalisable?
- Is the association likely to be causal?
- How important are the findings for Public Health?

Selection bias

- Can occur if selection of cases or controls is related to exposure of interest
  - eg study of smoking & lung cancer; controls with COPD
- Can occur if poor/differential response rates
  - Association between exposure and outcome may be different in those in the study vs those not included

Information bias: exposure data

- Reporting by cases and controls
  - Unreliable if exposure a long time ago
  - Differential (recall bias)
- Interviewing by observers
  - Probe more if aware of case-control status (and hypothesis)
- Minimise bias in exposure measurement by
  - Blinding of researchers to case control status
  - Blinding of participants to hypothesis

Importance of the prenatal environment

"The only clever thing I did was to remember that life begins at conception, not at birth…."

Alice Stewart

Prenatal X-rays and childhood malignancies

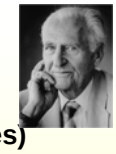
BMJ 1958; 1: 1495-1508

	Cases	Controls
X-rays		
Yes	141	81
No	1125	1204

OR = $\frac{141/1125}{81/1204} = \frac{0.125}{0.067} =$ **1.86**

Smoking and lung cancer

BMJ 1950; 739-48



Lung cancer (males)

	Cases	Controls
Smokers	647	622
Non-smokers	2	27

OR = $\frac{647/2}{622/27} =$ **14.0**

Cholera outbreak in Nigeria

J Water and Health 2003

		Cases	Controls
Drunk water from street vendors:	**Yes**	55	18
	No	44	55

OR = $\frac{55/44}{18/55} = \frac{1.25}{0.327} =$ **3.8 (1.9 – 7.9)**

Essential reading Week 6

- Relevant to this lecture (case control studies) although we won't discuss until **Week 7** seminar (Week 6 seminar relates to your assignment).
- Barker D, Cooper C, Rose G. Epidemiology in medical practice. Chapter 5.
- Fleming PJ et al. BMJ 1996; 313: 191-5.
- **NB** Please read this paper and the Introduction to the tutorial **BEFORE** the seminar in week 7.