How do we know that smoking causes lung cancer?

Seif Shaheen
Professor of Respiratory Epidemiology
Centre for Primary Care and Public Health
Blizard Institute
Barts and The London School of
Medicine and Dentistry

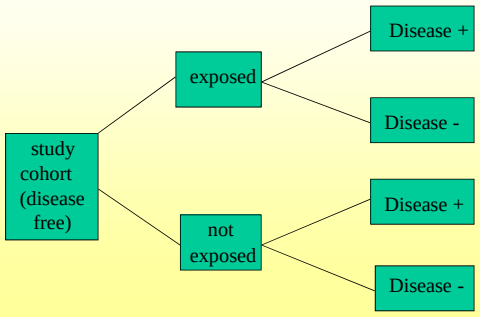
s.shaheen@qmul.ac.uk

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

Objectives

- Cohort studies
  - Prospective and historical
  - Strengths and weaknesses
- Measuring risk
  - Absolute risk, relative risk
  - Attributable risk, population attributable fraction
- Causality
  - Criteria for causal inference
  - Using genetic epidemiology

Prospective cohort study: design



Conducting a cohort study: five steps

- Select cohort population
- Measure exposure
- Follow-up
- Measure disease outcome
- Estimate disease risk associated with exposure

Selection of exposed and non-exposed groups

- Common exposures eg smoking, diet
  - General population cohort
  - Internal comparisons of exposure status
- Rare exposures
  - Cohort defined by geography, environmental exposure/disaster
    - Montserrat volcano
  - Cohort defined by occupation eg asbestos workers
    - Internal comparison (other workers in same industry)
    - External comparison (workers in different industry)

Measuring exposure to risk factors

- Records
  - Hospital eg birth weight
  - Occupational eg dust exposure
- Environmental monitoring
  - dust mite, NO_2 levels in air
- Lifestyle questionnaire
  - smoking, diet, occupation
- Clinical/biochemical/molecular measurement
  - Body Mass Index, nutrient biomarker, genotype

Follow-up

- A challenge!
  - Chronic diseases have long latent period
- Optimising follow-up
  - stable population eg Isle of Wight, Framingham
  - motivated population eg health personnel
  - regular contact and tracing
  - important to minimise BIAS

Measuring outcome

- Records
  - Mortality
    - Death certificates
  - morbidity
    - Health care records
- Interview / examination
  - questionnaire (standardised / validated)
    - chronic bronchitis
    - asthma
  - clinical/biochemical
    - lung function, blood pressure, blood sugar

Categorising exposure for analysis

- "Natural" categorical variable
  - Smoker
    - Yes/No
    - Never/Ex/Current
- Categorical variable from continuous variable
  - body mass index
    - <20; 20-24.99; 25-29.99; ≥30
    - Quantiles
- More than two categories is more informative
  - "Dose-response"

Defining outcome for analysis

- Binary outcome: Yes/No
  - Death; asthma
    - Analyse risk
- Continuous outcomes eg lung function, bp
  - Define "disease" (Yes/No) using cut-off
    - Eg COPD: $FEV_1/FVC<70\%$
    - Analyse risk
  - Keep continuous outcome
    - Analyse difference in mean outcome between exposure groups

Comparing disease risk in exposed and non-exposed (1)

- Count number of new cases of disease in each exposure group
- Risk (incidence) of disease
  = number of new cases during defined period / total number at risk at start of period
- Relative risk (risk ratio) = risk in exposed / risk in non-exposed

Calculating the relative risk

	Develop disease		
	Yes	**No**	**Total**
Exposed			
Yes	a	b	a+b
No	c	d	c+d

Relative risk = $\frac{a/(a+b)}{c/(c+d)}$

Obesity and adult-onset asthma

- Nurses Health Study, USA
- 85, 911 participants aged 26-46 in 1991
- Body Mass index measured at baseline
- Followed up for 4 years
- Outcome measure: incident asthma

Arch Intern Med 1999; 159: 2582-8

Obesity and adult-onset asthma

	Develop asthma		
	Yes	**No**	**Total**
Obese			
Yes (BMI ≥30)	398	10,805	11,203
No	1,198	73,510	74,708

Relative risk = $\frac{398/11,203}{1,198/74,708}$ = **2.22 (1.98 - 2.48)**

Obesity and adult-onset asthma

BMI	**Adjusted RR (95% CI)**
< 20	0.9 (0.7 - 1.1)
20-22.4	1.00
22.5-24.9	1.1 (1.0 - 1.3)
25.0-27.4	1.6 (1.3 - 1.9)
27.5-29.9	1.7 (1.4 - 2.0)
≥30	2.7 (2.3 - 3.1)

Paracetamol and adult-onset asthma: Nurses Health Study

Frequency of use (days per month)	*Adjusted RR (95% CI)*	
None	1.0	
1-4	1.27	(0.96 to 1.66)
5-14	1.43	(0.99 to 2.07)
15-21	1.78	(1.04 to 3.04)
22+	1.53	(0.95 to 2.46)

P trend = 0.006

Leisure time physical activity and risk of death

- Copenhagen City Heart Study, Denmark
- 7,023 men and women aged 20-79
- Physical activity measured in 1976-8 and 1981-3
- 1424 men and 1301 women died during 17-year follow-up

AJE 2003; 158: 639-44

Physical activity and mortality risk

Level of activity at 2nd exam in those who had low activity at first exam	Men RR* (95% CI)	Women RR* (95% CI)
Low	1.00	1.00
Moderate/high	0.64 (0.50, 0.81)	0.74 (0.58, 0.95)

*Adjusted relative risk

Comparing disease risk in exposed and non-exposed (2)

- **Rate**

  = number of new cases during defined period
  total "person time at risk" during period

- **Relative rate** (rate ratio) = rate in exposed
  rate in non-exposed

Nut consumption and CHD

- Nurses Health Study, USA
- 1980-1994
  - 1,132, 229 person years of follow-up
- Dietary questionnaire at baseline
  - nut consumption
- 1255 new cases of coronary heart disease

Hu et al, BMJ 1998; 317: 1341-45

Nut consumption and CHD

Freq of eating nuts	**Cases**	**Person years**	**RR* (95% CI)**
Never	542	391,918	1.0
<2x/week	584	579,805	0.91 (0.81,1.03)
2-4x/week	85	102,175	0.78 (0.61,0.99)
≥5x/week	44	58,330	0.66 (0.47,0.93)

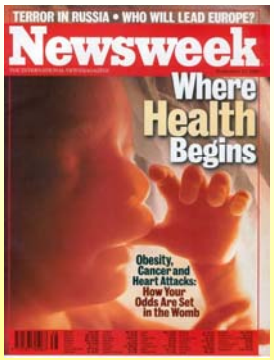
*adjusted Relative Rate

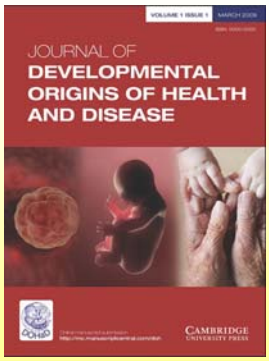
P trend 0.005

Prospective cohort studies

- Strengths
  - study rare exposure
  - study multiple effects of one exposure
  - demonstrate temporality
  - minimise bias in exposure measurement
  - measure incidence
- Limitations
  - inefficient for rare diseases
  - costly and time-consuming
  - potential bias from losses to follow-up

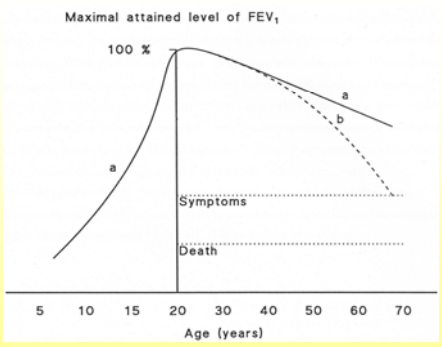
A revolutionary new idea….



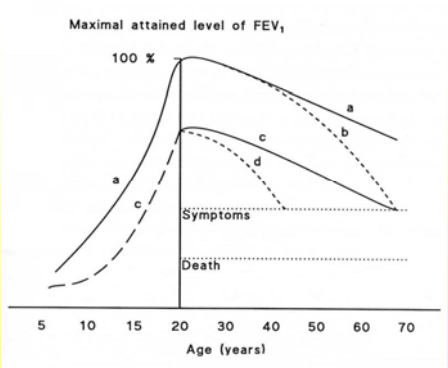


Historical cohort studies

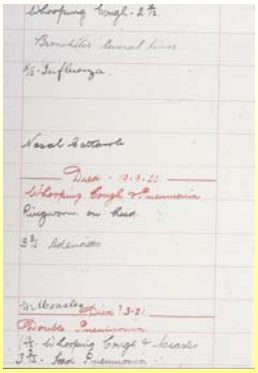
- How do they differ from prospective cohort studies?
  - Outcome of interest has already occurred when study begins, therefore efficient for diseases with long latent periods
- How do they differ from case control studies?
  - Individuals selected according to documented exposure status (historical records)



Speizer and Tager, Epidemiol Rev 1979;1: 124-42.



South Derbyshire: born 1917-1922

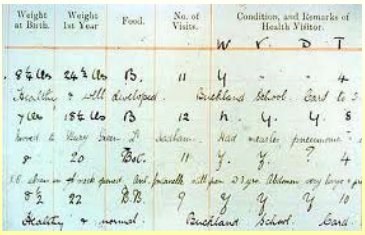


Mean FEV_1 (l), adj. for age and height, among D'shire men and women aged 67-74 (n=618)

	Mean FEV_1			
	Pneumonia <2yrs		**Diff in FEV_1 (95% CI)**	
	No	*Yes*		
Men	2.35	1.69	-0.65	(-1.02, -0.29)
(n)	(315)	(13)		
Women	1.70	1.52	-0.19	(-0.51, +0.14)
(n)	(279)	(7)		

Am J Respir Crit Care Med 1995; 151:1649-52

Hertfordshire: born 1920-1930

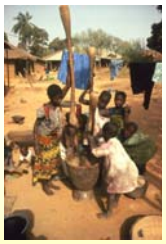


Mean FEV_1 (l), adjusted for age and height, among Herts men aged 59-67 (n=639)

	Infant bronchitis/pneumonia	
Birth wt (lbs)	*Absent*	*Present*
≤5.5	2.39 (22)	1.81 (4)
-6.5	2.40 (70)	2.23 (10)
-7.5	2.47 (163)	2.38 (25)
-8.5	2.53 (179)	2.33 (12)
-9.5	2.54 (103)	2.36 (5)
>9.5	2.57 (43)	2.36 (3)

BMJ 1991; 303:671-5

And so to Africa……



Guinea-Bissau, 1994; children who survived severe measles epidemic in 1979

Atopy according to measles history in 14-21year-olds in Bissau (n=262)

	Atopy %	OR* (95% CI)
Measles		
No (n=129)	25.6	1.0
Yes (n=133)	12.8	0.36 (0.17, 0.78)

* controlling for potential confounders

Lancet 1996; 347: 1792-6.

Why is it potentially misleading just to report the relative risk or odds ratio?

- The RR or OR only tells us about the aetiological, not public health, importance of an exposure
- The RR or OR alone may lead to “hype” by the media and unnecessary alarm for the public

How important are findings for Public Health?

- How many excess cases among exposed can be attributed to exposure?
  - Attributable risk
- What proportion of disease in the population can be attributed to exposure?
  - Population Attributable Fraction
- Gives an idea of scope for prevention if exposure removed (assuming causal relation)

Nurses' Health Study: Risk of primary PE by postmenopausal hormone use (1)

	Cases	Person-years	RR* (95% CI)
HRT use			
Never	27	320,339	1.0
Current	22	155,669	2.1 (1.2 to 3.8)

* adjusted relative rate

(Grodstein et al. Lancet 1996)

Nurses' Health Study: Risk of primary PE by postmenopausal hormone use (2)

HRT use	Cases	Person-years	Absolute rate
Never	27	320,339	8 /100,000 /yr
Current	22	155,669	14 /100,000 /yr

Attributable Risk = risk in exposed - risk in non-exposed
= 6 cases /100,000 women /year

The importance of reporting absolute and attributable risks

- Puts research findings (RRs and ORs) into perspective
  - For policy makers
    - Do we need to do anything about this risk factor?
    - "all policy decisions should be based on absolute measures of risk; relative risk is strictly for researchers only" (Geoffrey Rose, 1991)
  - For the public
    - Should we be worried about this risk factor?
- Enables fuller interpretation and better communication of risk

Does asthma begin in utero?

- Early presentation

- Prenatal risk factors
  - Maternal smoking in pregnancy
  - Antibiotic use in pregnancy
  - Infections in pregnancy
  - Complications of pregnancy
  - Mode of delivery
  - Gestational age at birth
  - Anthropometry at birth

Avon Longitudinal Study of Parents and Children (ALSPAC)

- Prospective study of 14,541 pregnancies
  - 14,062 live births
  - 13,988 survived to 1 year
- Eligible
  - EDD 1.4.91 - 31.12.92
  - resident in Bristol health districts
- Enrolled
  - as early as possible in pregnancy
  - 85-90% of those eligible

Data collected

- Maternal and child questionnaires
  - Prenatal nutrition
    - Biomarkers
    - FFQ in late pregnancy
  - Other prenatal/childhood exposures/confounders
- DNA on 10,000 mothers and 10,000 children
- Respiratory and atopic phenotypes
  - Early childhood wheezing phenotypes
  - Asthma, wheezing and atopic disease at 6 years
  - Skin test reactivity and total IgE at 7 years
  - Lung function and BR (methacholine) at 8-9 years

Paracetamol use in late pregnancy and prevalence of wheezing at 30-42 months

Frequency	n	%
Never	608/5134	11.8
Some days	561/3725	15.1
Most days/daily	26/88	29.5

Paracetamol use in late pregnancy and risk of wheezing at 30-42 months

(Thorax 2002; 57: 958-63)

Frequency	OR (95% CI)	Adj OR (95% CI)
Never	1.00	1.00
Some days	1.34 (1.18, 1.52)	1.12 (0.98, 1.28)
Most days/daily	3.17 (1.99, 5.05)	2.10 (1.30, 3.41)*

*P=0.003

Paracetamol in pregnancy and childhood wheezing: Interpretation


MIRROR (LONDON, UK) 30th October 2002

PREGNANT MUMS USING PAINKILLERS DOUBLE RISK OF ASTHMA IN BABIES DOC WARNS OF LINK

- BUT
  - Population Attributable Fraction = ~1%

Causal inference in observational studies

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

Strengthening causal inference

- Gene by environment interaction
  - Modification of paracetamol effect by gene variants influencing toxicity: ↑ bio plausibility
    - nb human data lacking
- Glutathione-S-transferase
  - GSTT1, GSTM1, GSTP1
  - conjugates NAPQI with GSH
- Nrf2
  - Knockout mice sensitive to paracetamol toxicity
  - Disruption of Nrf2 leads to increased allergic inflammation in a mouse model of asthma

Paracetamol use in early pregnancy and asthma risk stratified by maternal Nrf2

	Adj OR*	**95% CI**	**P**
C:C (n=3754)	0.99	0.81 to 1.21	0.91
T:C/T:T (n=1137)	1.73	1.22 to 2.45	0.002
		Interaction	*0.02*

**Per category of exposure*

No interaction with child Nrf2 genotype

Risk of impaired lung function by maternal smoking and GSTM1

Overall: -0.043* (-0.069 to -0.016); P trend 0.0017

Child genotype

GSTM1 present: -0.017 (-0.06 to 0.027)

GSTM1 null: -0.061 (-0.10 to -0.02)

Maternal genotype

GSTM1 present: -0.019 (-0.07 to 0.033)

GSTM1 null: -0.054 (-0.10 to -0.005)

*Age/height-adjusted deficit (95% CI) in FEF_{25-75} (SDs) associated with smoking (per category increase)

Quiz!

Which study design would be optimal in order to study the following?:

	Case-control	Cohort
Rare disease		
Rare exposure		
Multiple exposures		
Multiple outcomes		
Natural history of disease		
Disease rate		

Quiz answers

Which study design would be optimal in order to study the following?:

	Case-control	Cohort
Rare disease	✓	x
Rare exposure	x	✓
Multiple exposures	✓	**(✓)**
Multiple outcomes	x	✓
Natural history of disease	x	✓
Disease rate	x	✓

Bored to death……..?

IJE 2010

- Follow-up of Whitehall civil servants
  - Higher cardiovascular mortality in those reporting ‘a great deal’ of boredom at baseline compared with those who were ‘not bored at all’

Essential reading Week 7

- Relevant to this lecture (cohort studies)
- Barker D, Cooper C, Rose G. Epidemiology in medical practice. Chapter 5.
- Doll R, Peto R. BMJ 1976; 2: 1525-36.
- (Doll R. Am J Respir Crit Care Med 2000; 162: 4-6.)
- **NB** Please read the above and the Introduction to the tutorial **BEFORE** the seminar.