Introduction to epidemiological study design

Study = basic tool in epidemiology

"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

Epidemiology = comparison

- 550 cases of stomach cancer

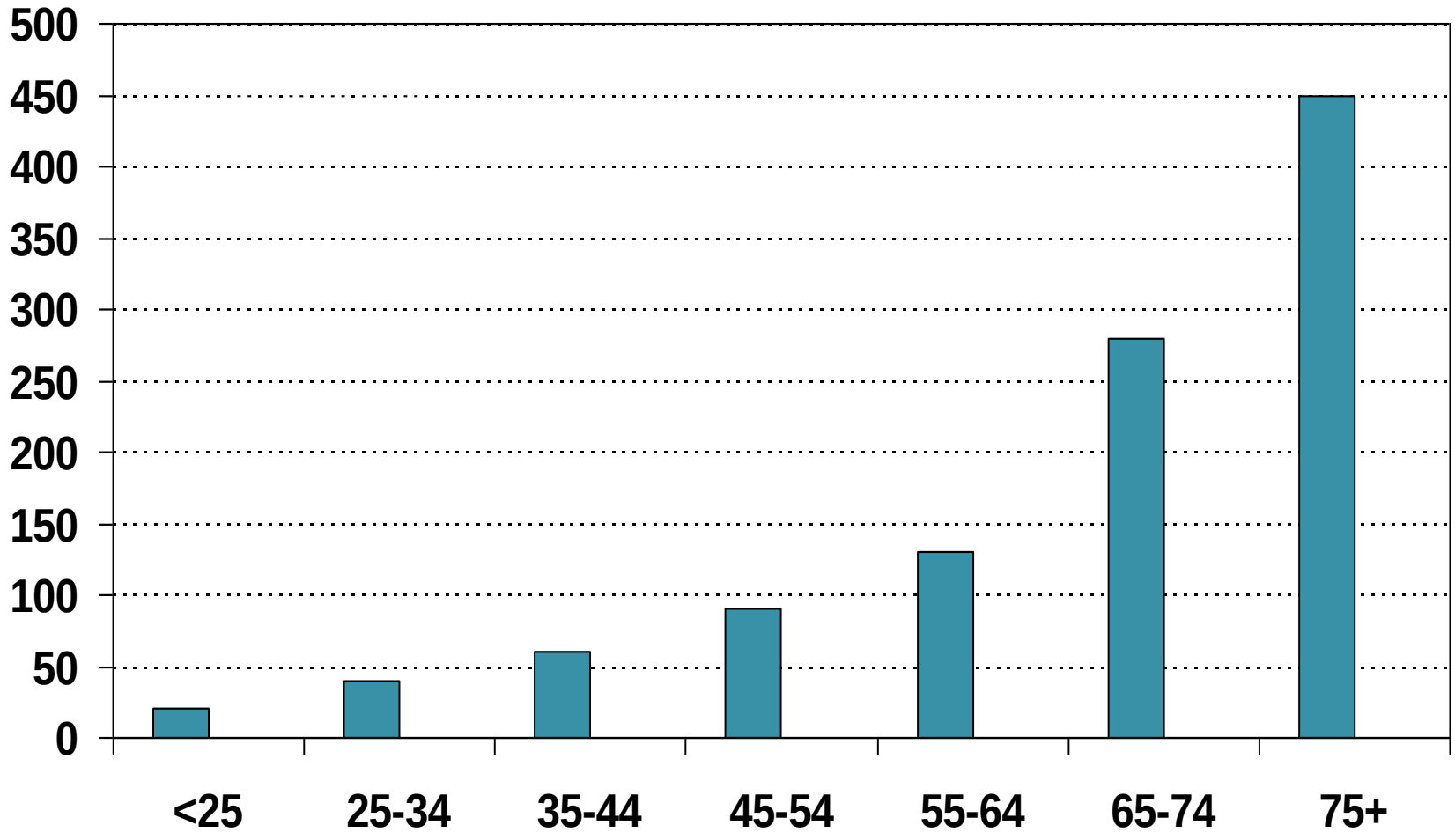
Epidemiology = comparison

- 550 cases of stomach cancer in Hertfordshire in 2005

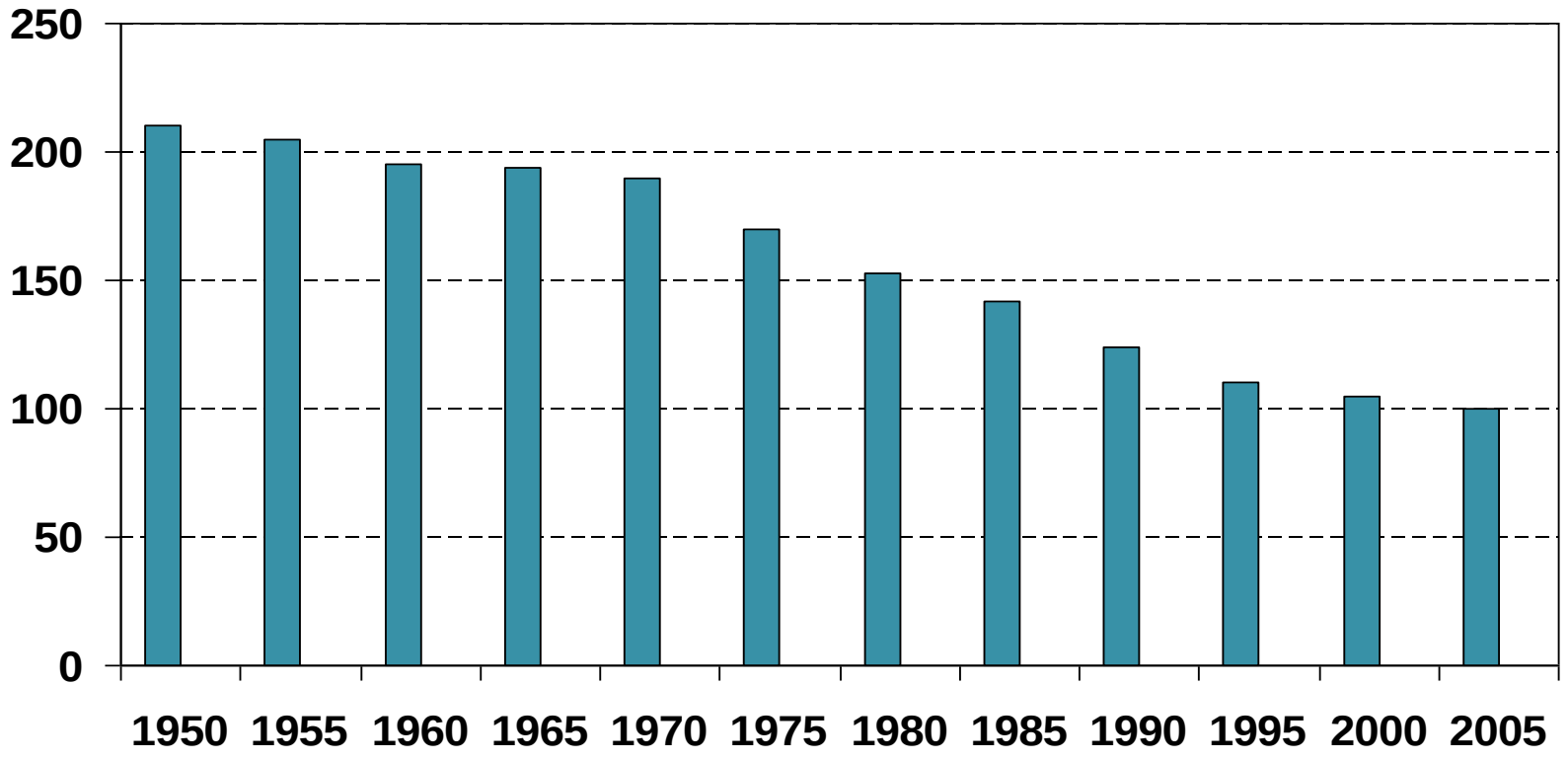
Epidemiology = comparison

- 550 cases of stomach cancer in Hertfordshire in 2005
- Population 550,000
- Rate 100/100,000

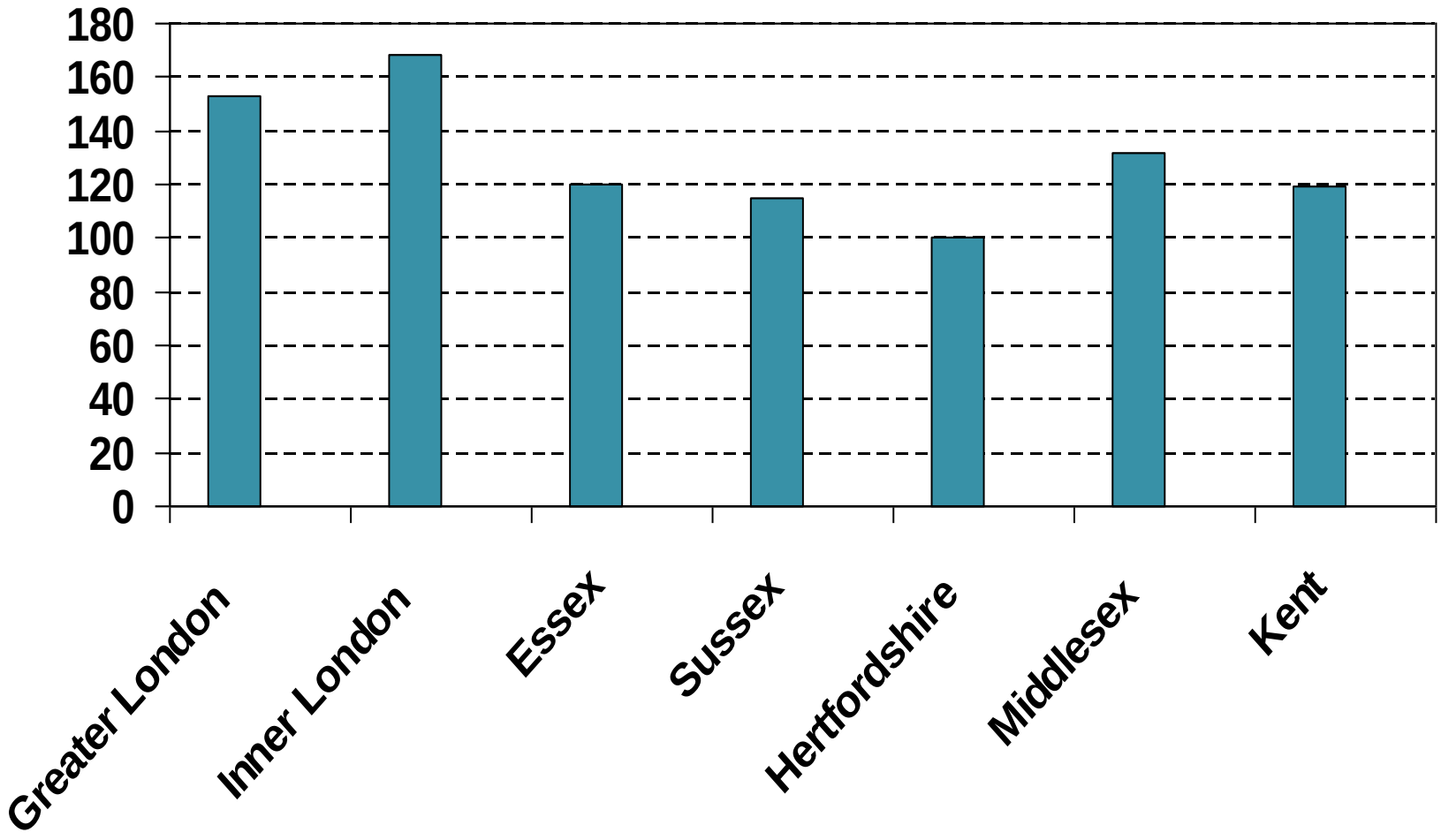
Stomach cancer by age group, 2005, per 100,000

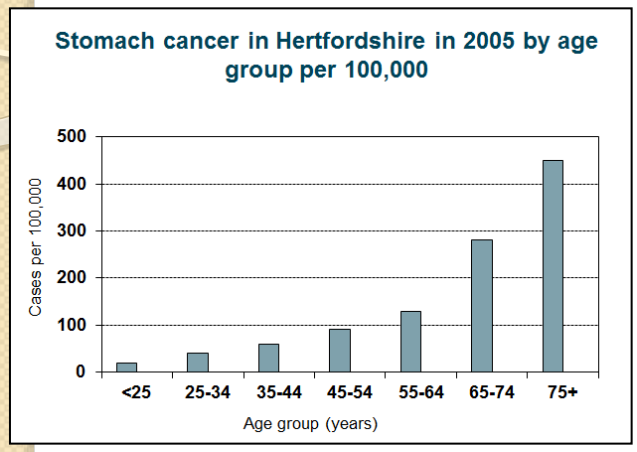


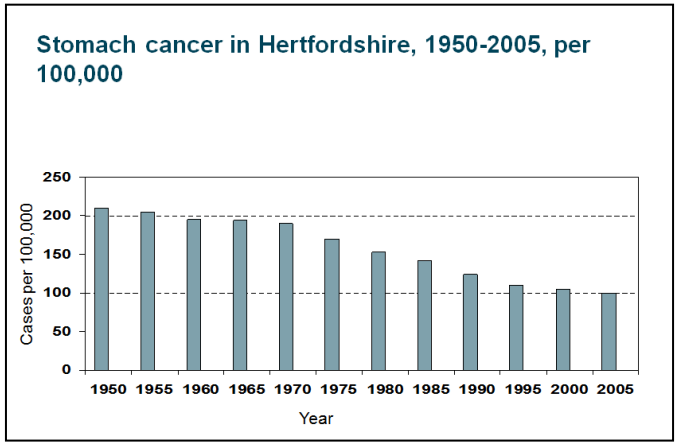
Stomach cancer in Hertfordshire, 1950-2005, per 100,000

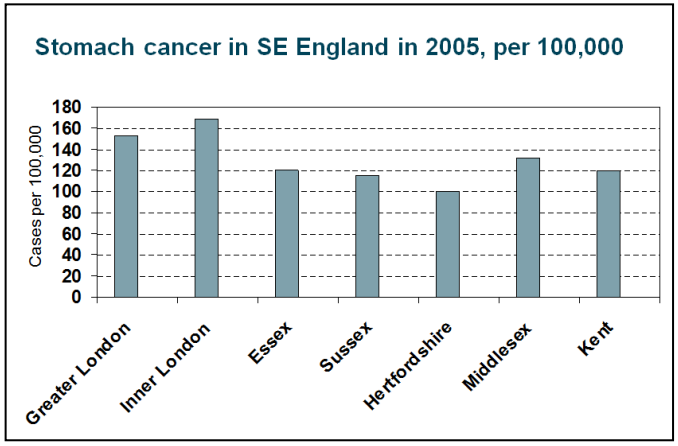


Stomach cancer in SE England in 2005, per 100,000




Stomach cancer in Hertfordshire in 2005 by age group per 100,000
Cases per 100,000
500
400
300
200
100
0
<25
25-34
35-44
45-54
55-64
65-74
75+
Age group (years)

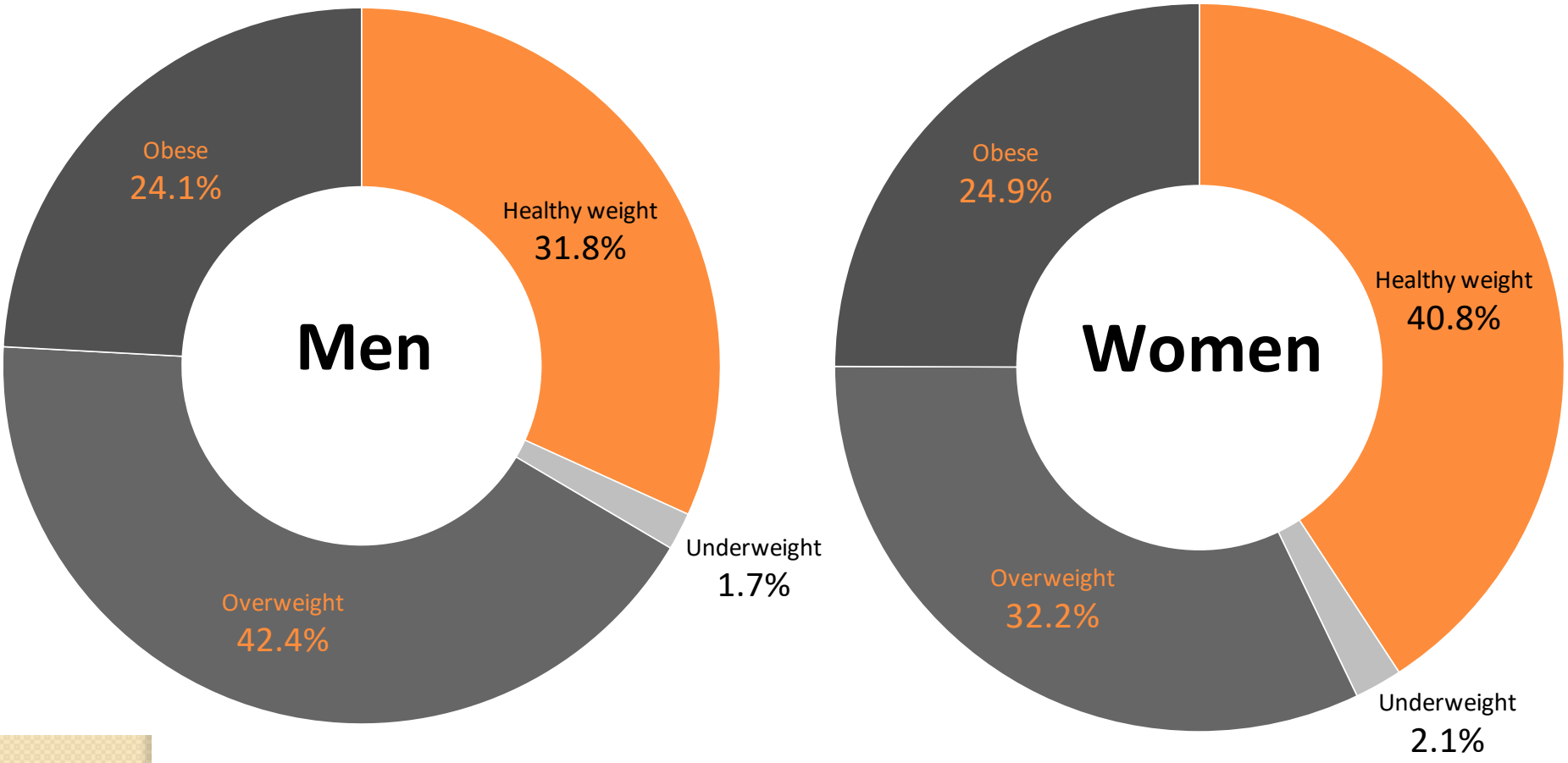

Stomach cancer in Hertfordshire, 1950-2005, per 100,000
Cases per 100,000
250
200
150
100
50
0
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
Year


Stomach cancer in SE England in 2005, per 100,000
Cases per 100,000
180
160
140
120
100
80
60
40
20
0
Greater London
Inner London
Essex
Sussex
Hertfordshire
Middlesex
Kent

Another example

Adult prevalence by BMI status

Health Survey for England (2008-2010 average)



Adult (aged 16+) BMI thresholds

Underweight: <18.5kg/m^2

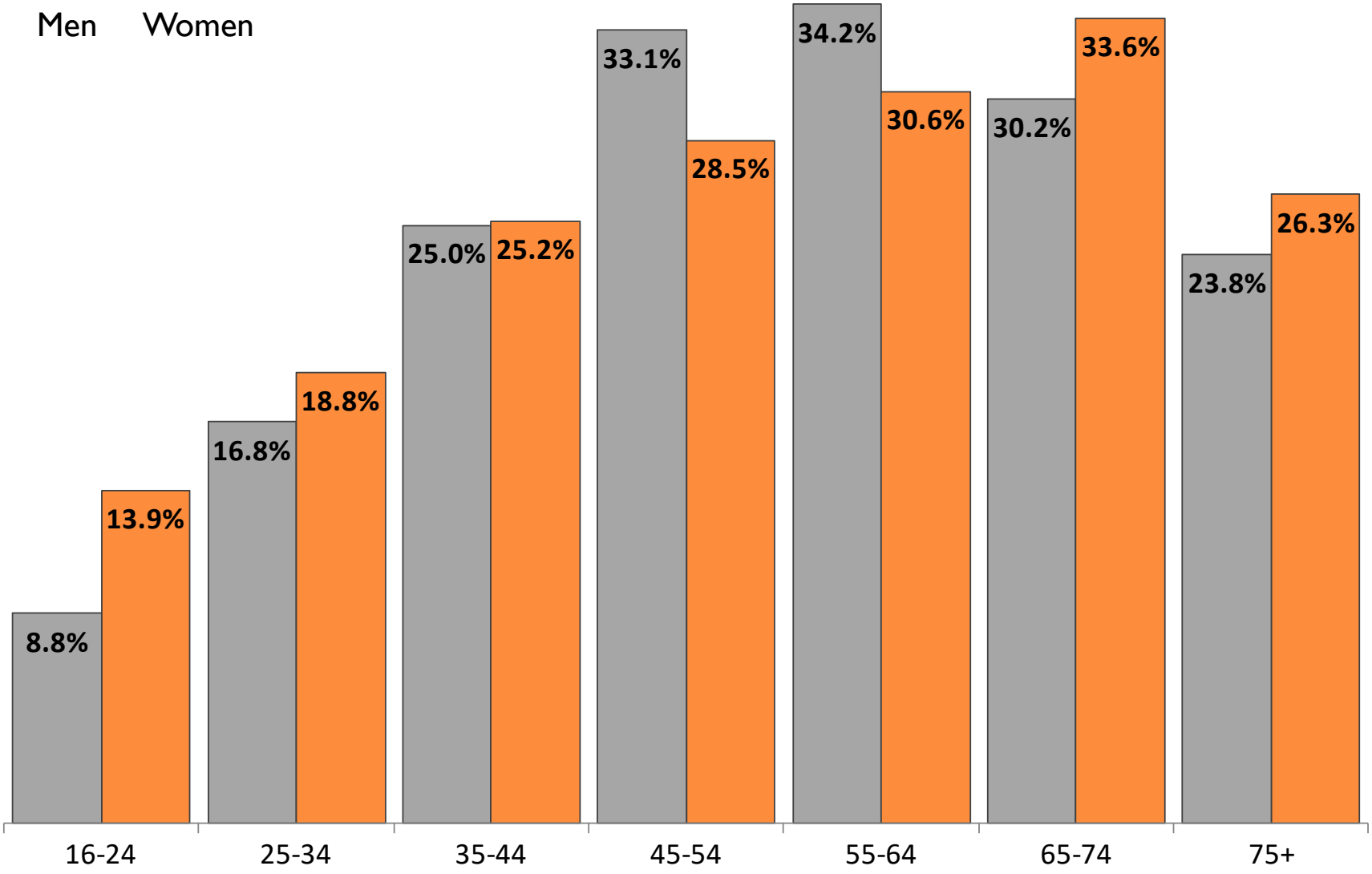
Healthy weight: 18.5 to <25kg/m^2

Overweight: 25 to <30kg/m^2

Obese: ≥30kg/m^2

Adult obesity prevalence by age and sex

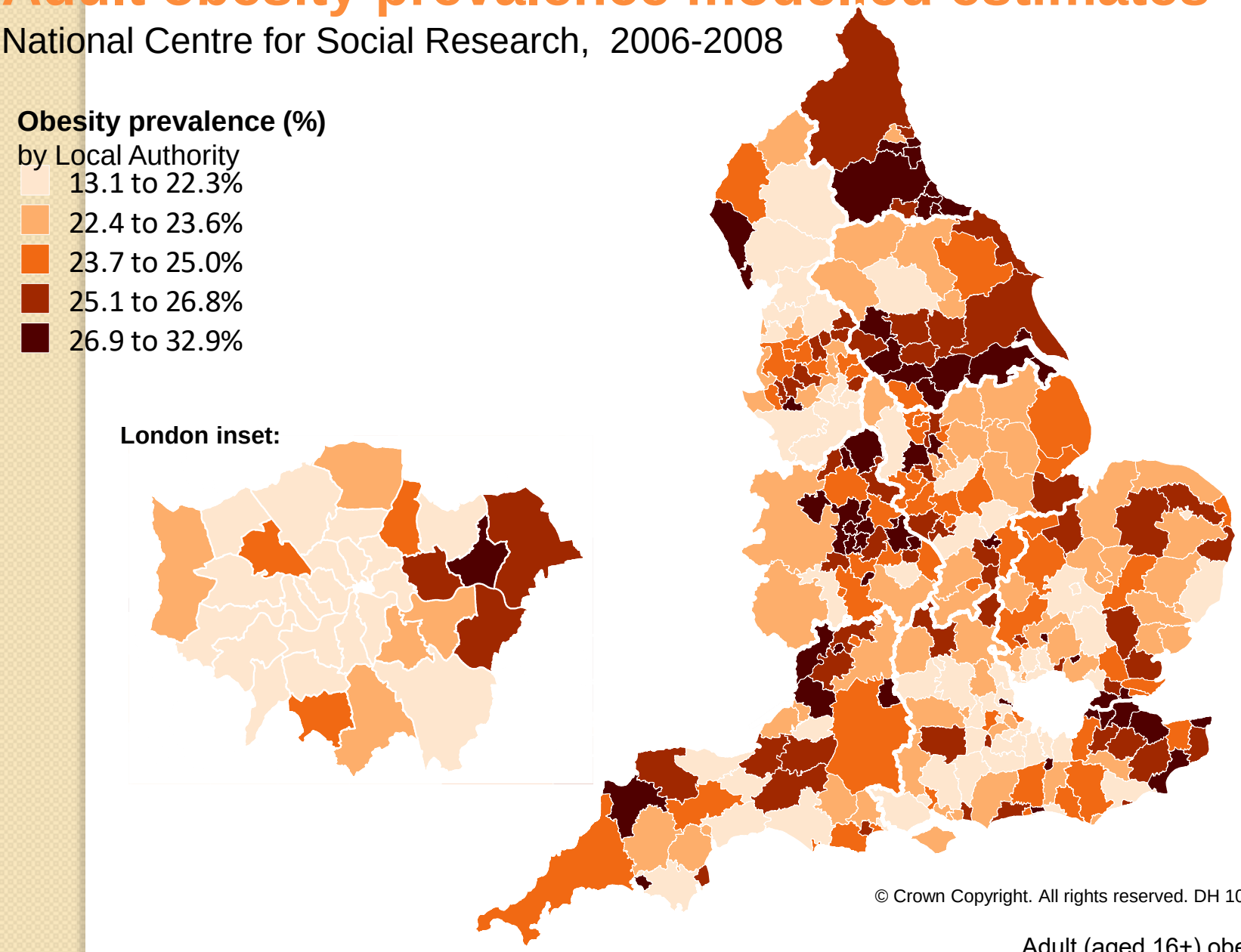
Health Survey for England 2008-2010



Adult (aged 16+) obesity: BMI ≥ 30kg/m^2

Adult obesity prevalence modelled estimates

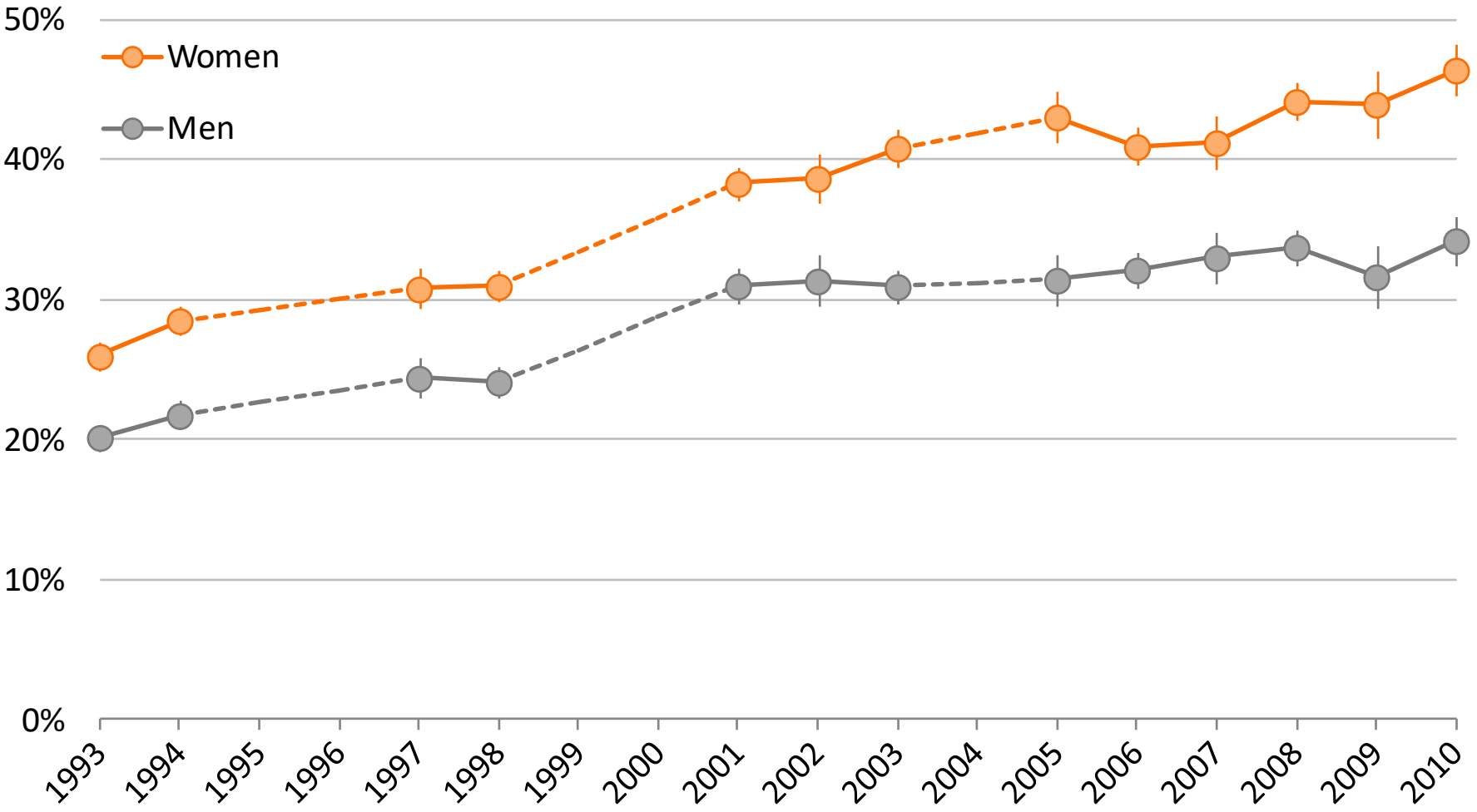
National Centre for Social Research, 2006-2008



Adult (aged 16+) obesity: BMI ≥

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

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

Two primary criteria

- Descriptive vs. analytical
- Observational vs. interventional

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)

Descriptive studies

- Describe patterns of disease occurrence
- 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

Descriptive studies
4 Ws : What? Who? Where? When?

What? ……. health outcome / case / event

- Mortality
- Dental health
- Chronic disease
- Cognitive function

Descriptive studies
4 Ws : What? Who? Where? When?

Person (Who?)

Age, sex, marital status, social class ….

Place (Where?)

Regions (disease atlases), internationally (Japan vs. USA)

Time (When?)

When events occurred:

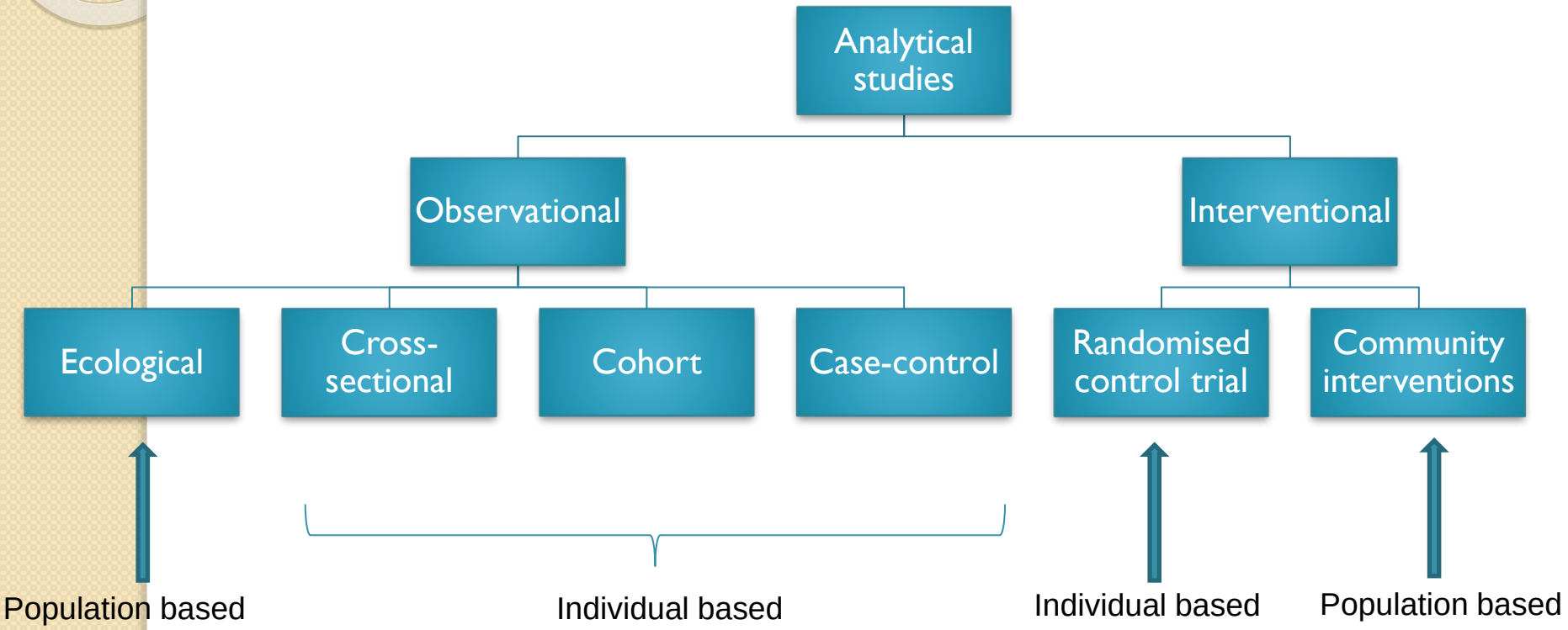
- sudden onset of diseases
- seasonal pattern (births, deaths, infections)
- secular trends

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

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)

Analytical studies

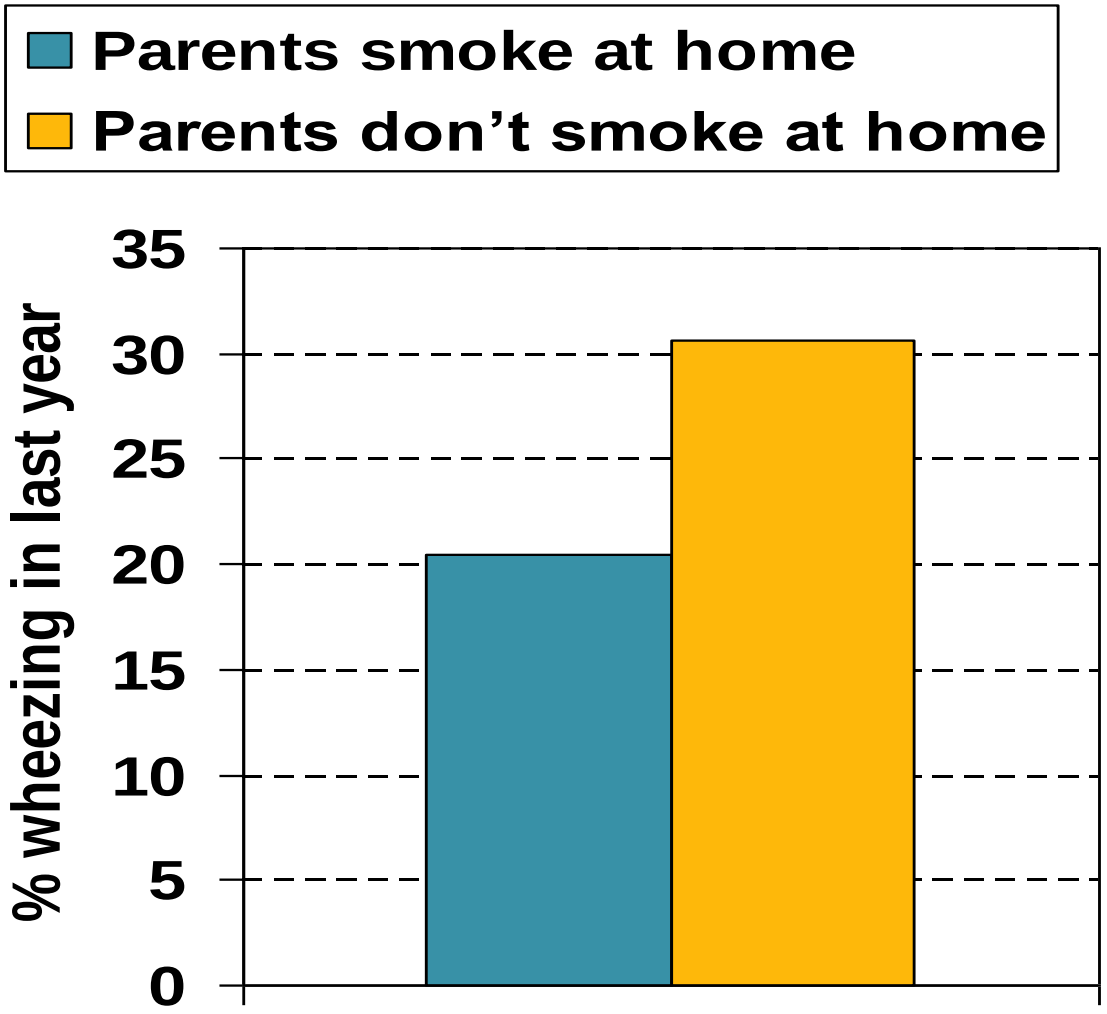


Observational vs. interventional studies

- ***Observational studies*** are studies which observe the populations or individuals under study; they normally include:
  - descriptive studies
  - ecological studies
  - cross-sectional studies
  - cohort studies
  - case-control studies
- ***Interventional studies*** are those where the investigators intervene, e.g. they assign exposure or a health measure to a particular individuals or groups. They include:
  - Prevention studies
  - Randomised clinical trials
  - Community interventions

Cross-sectional studies

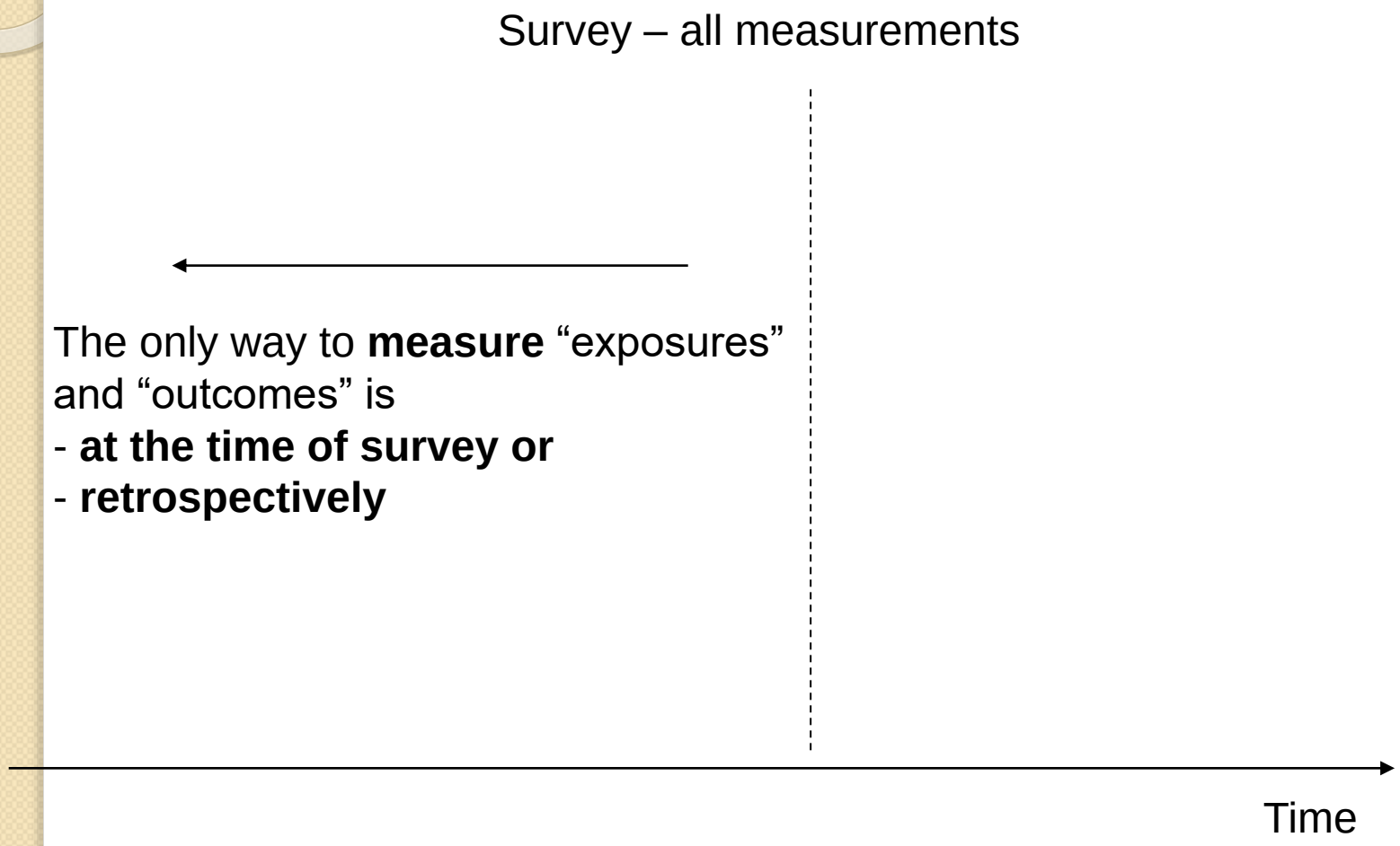
Example


Parents smoke at home
Parents don't smoke at home
% wheezing in last year
35
30
25
20
15
10
5
0

Cross-sectional studies

- In a cross-sectional study, all information is **collected at one point in time**
  - Outcome
  - Exposures
  - Covariates
- Sometimes called "survey"
- Cross-sectional studies could be descriptive or analytical
- **Always observational**
- The unit of analysis is the individual

Cross-sectional study

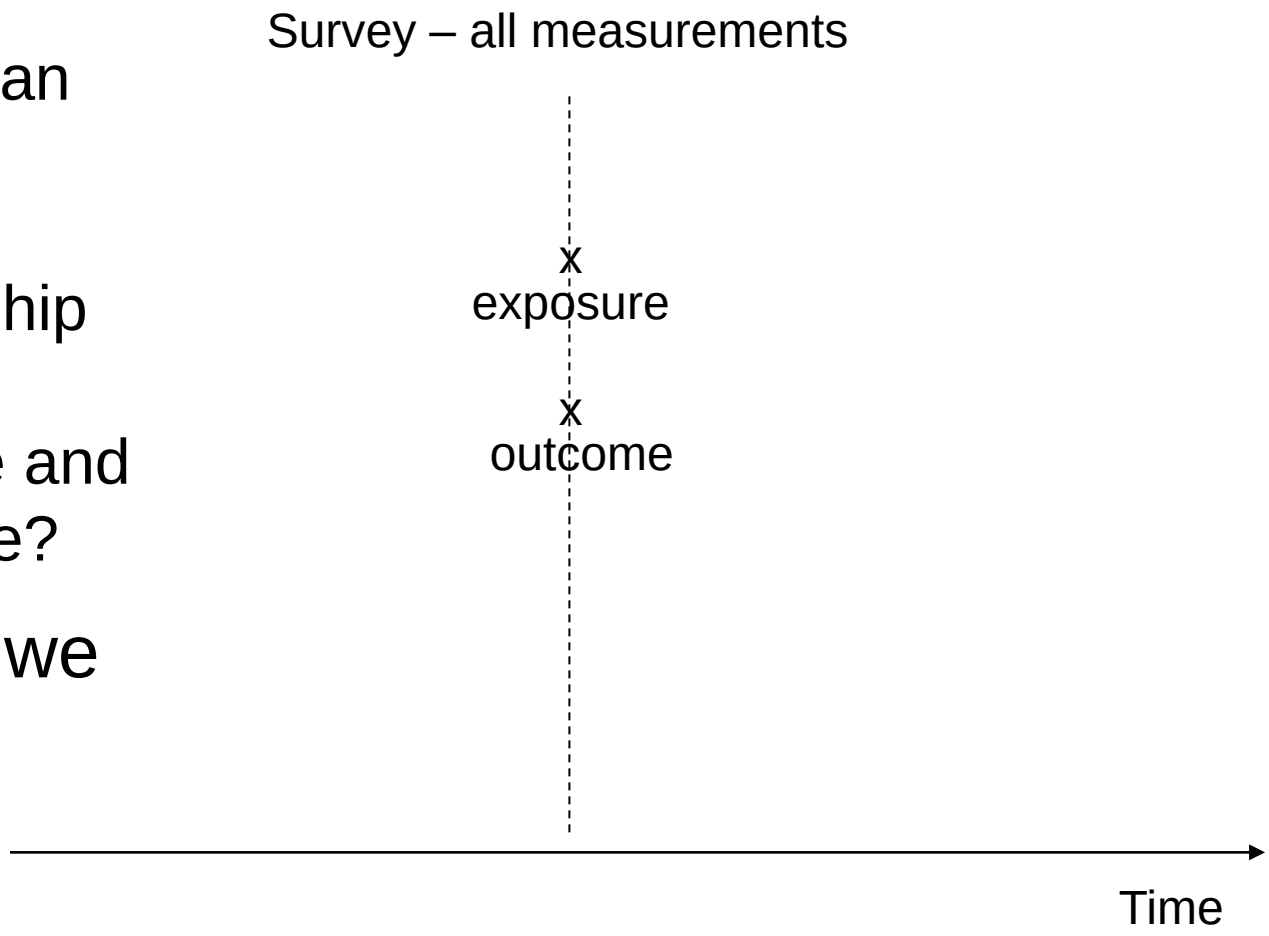


Cross-sectional studies: Advantages

- Relatively quick, do not require follow up
- Provide a snapshot, e.g. prevalence of a disease or a risk factor in population
- Allow examination of multiple diseases and multiple exposures
- Can test or suggest hypotheses

Cross-sectional studies: Limitations

- What can we say about relationship between outcome and exposure?
- What can we not say?



Cross-sectional studies: Limitations

- Since both disease and exposures are measured at the same time, **temporality is unclear**
- Difficult to estimate past exposure, especially if it occurred long time ago. **Not ideal for studying exposures that change over time** (e.g. diet). (but no problem with factors that are stable over time, e.g. genetic markers.)
- **Sensitive to reporting or recall bias** if exposures are subjectively reported.
- **Sensitive to response rates and representativeness** if used to estimate prevalence of a condition in population.

Representativeness

- Cross-sectional studies are often used to estimate the frequency of a condition in a population but it is usually impossible to study the whole population
- The validity of such estimates depends critically on the representativeness of the studied sample
- Response rate also important

What if…

75% response rate, and prevalence of 25% in responders

Prevalence in non-responders	Total prevalence (in full sample)
0	19%
25%	25%
50%	31%
75%	38%
100%	44%

Ecological studies

Ecological studies

- The unit of analysis is **a group** (e.g. country, district, population etc)
- Data cannot be disaggregated to the level of an individual.
- Also sometimes called *correlation* studies or *geographical* studies
- Include comparisons over time (time-series)
- Usually cheap and quick

Fish consumption and mortality

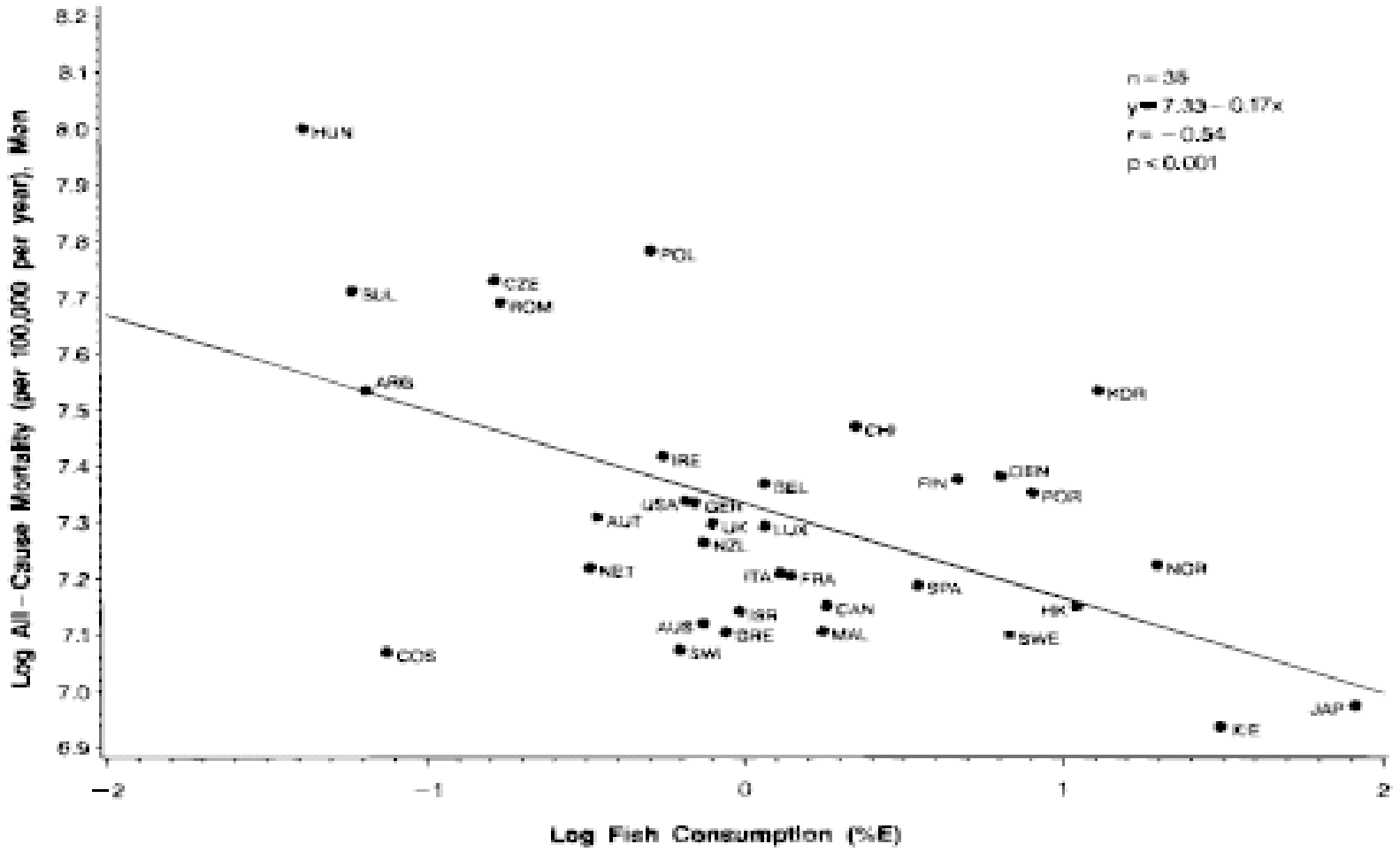
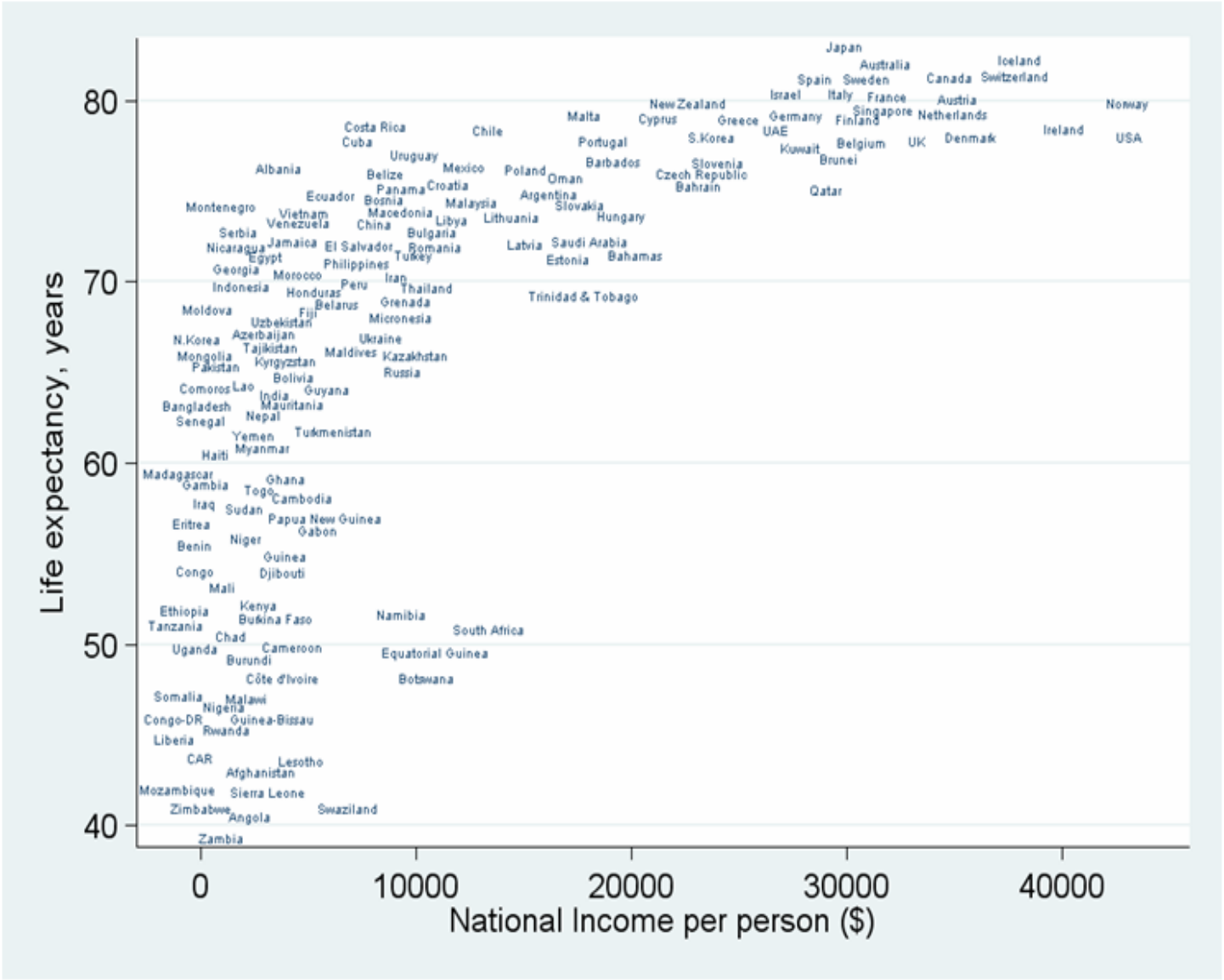


FIG. 1. Univariate relation between log fish consumption (%E, percent of total energy), 1989–1991 average and log all-cause mortality rate, age-standardized to 45–74 years, mean of the latest available 3 years in men. ARG, Argentina; AUS, Australia; AUT, Austria; BEL,

Income per head and life-expectancy: rich & poor countries



Source: Wilkinson & Pickett, The Spirit Level (2009)

Ecological fallacy

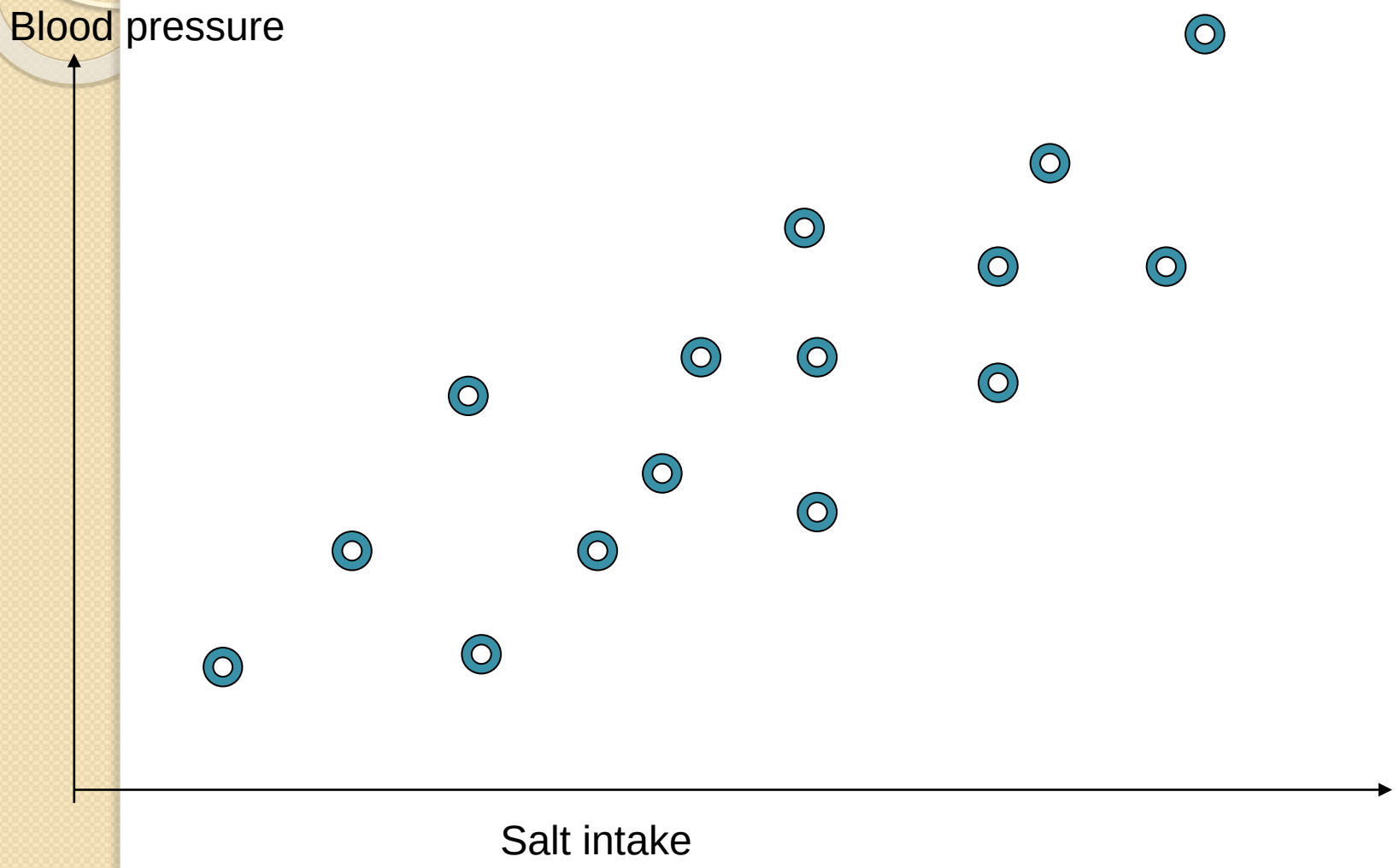
- This is a logical fallacy in the **interpretation** of statistical data where inferences about the nature of individuals are deduced from inference for the group to which those individuals belong
- **Extrapolation from groups to individuals** is conceptually inappropriate
- Situation when individual-level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

Ecological fallacy – example

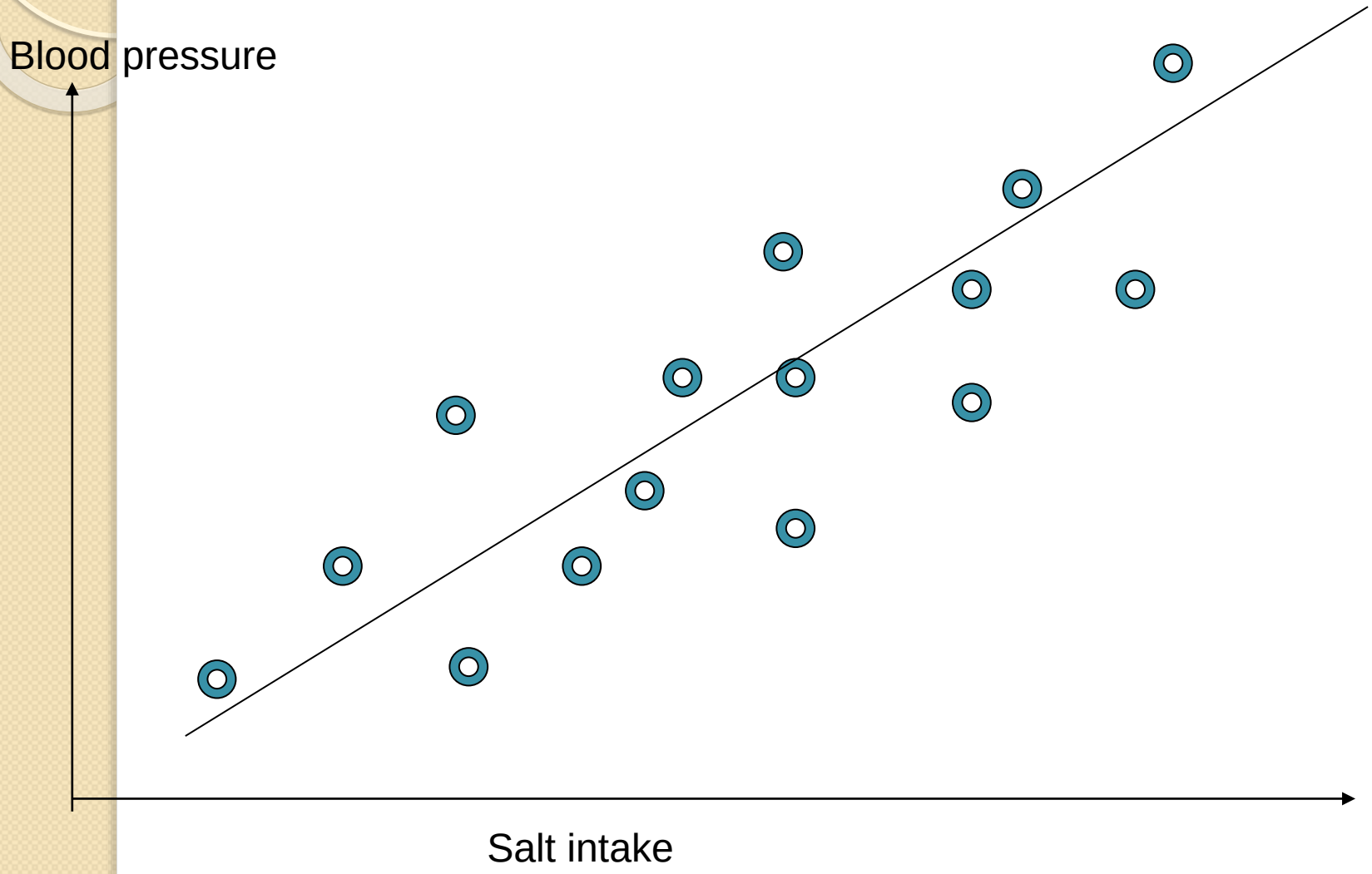
- Illiteracy rate and the proportion of the population born outside the US:
- State-level correlation: -0.53 (the higher % of immigrants, the lower the state's average illiteracy)
- Individual-level correlation: +0.12 (immigrants were on average more illiterate than native citizens)
- Immigrants tended to settle in states where the native population was more literate.

Robinson, W.S. (1950). "Ecological Correlations and the Behavior of Individuals". American Sociological Review; 15 (3): 351–357.

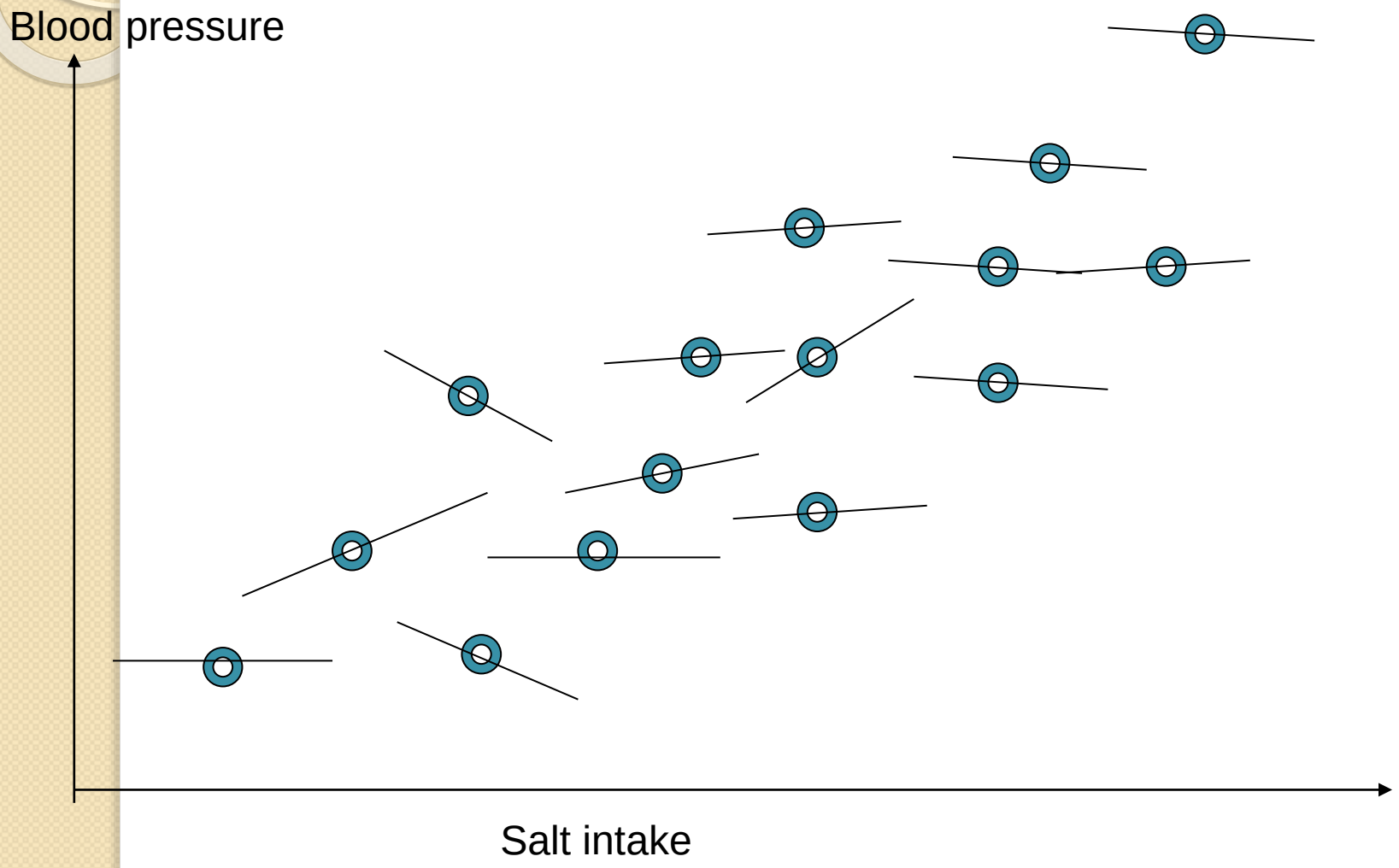
Ecological fallacy (1)



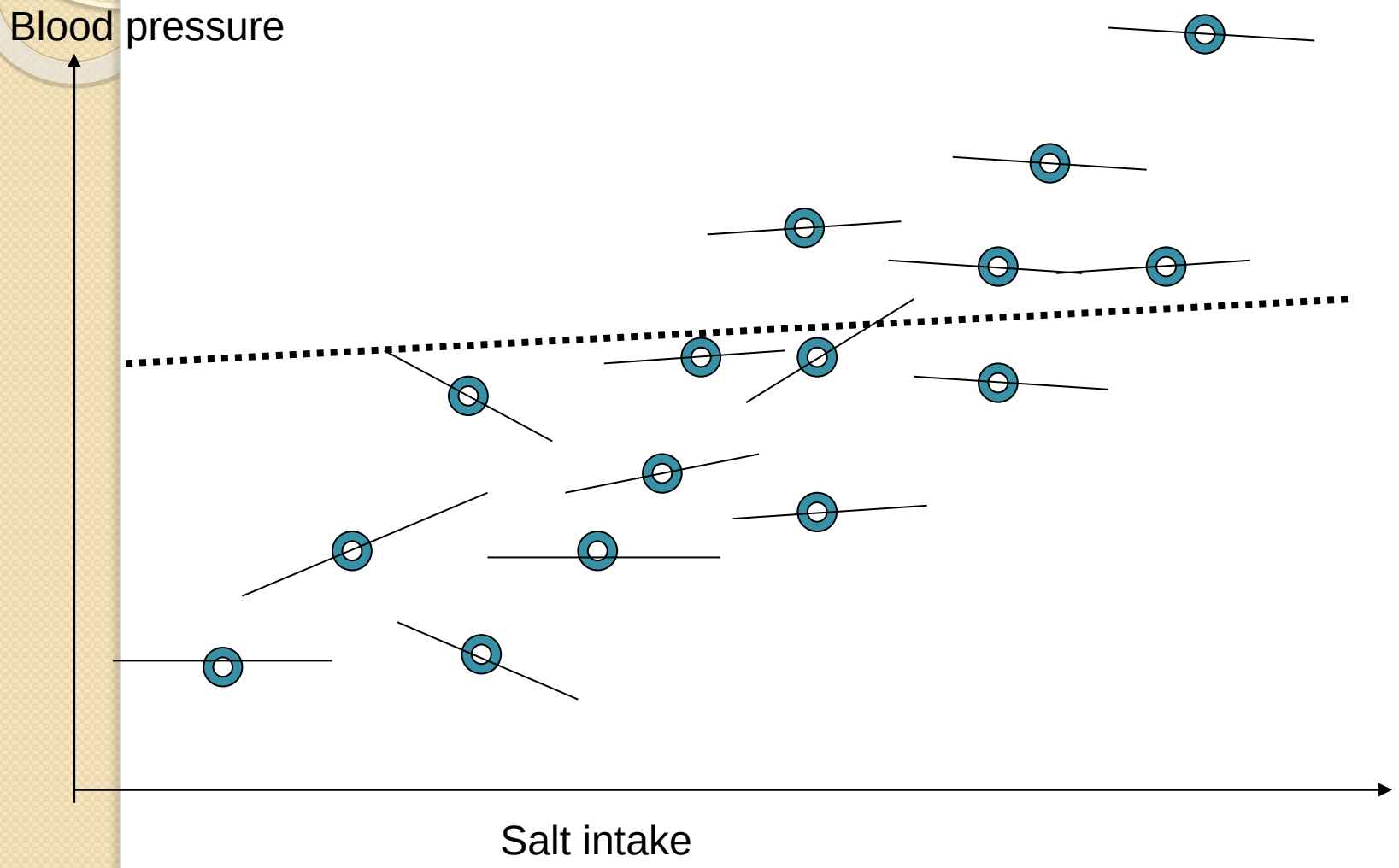
Ecological fallacy (2)



Ecological fallacy (3)



Ecological fallacy (4)



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

Time-series studies

- Studies repeated over time
- But not on the same individuals (i.e. not longitudinal)
- Type of ecological studies because subjects / events / exposures are grouped by a time interval, hard to disaggregate individuals
- For example, health survey on a representative sample repeated every 10 years… individual data collected but not on the same individuals at each survey
- They are useful for comparing changes over time

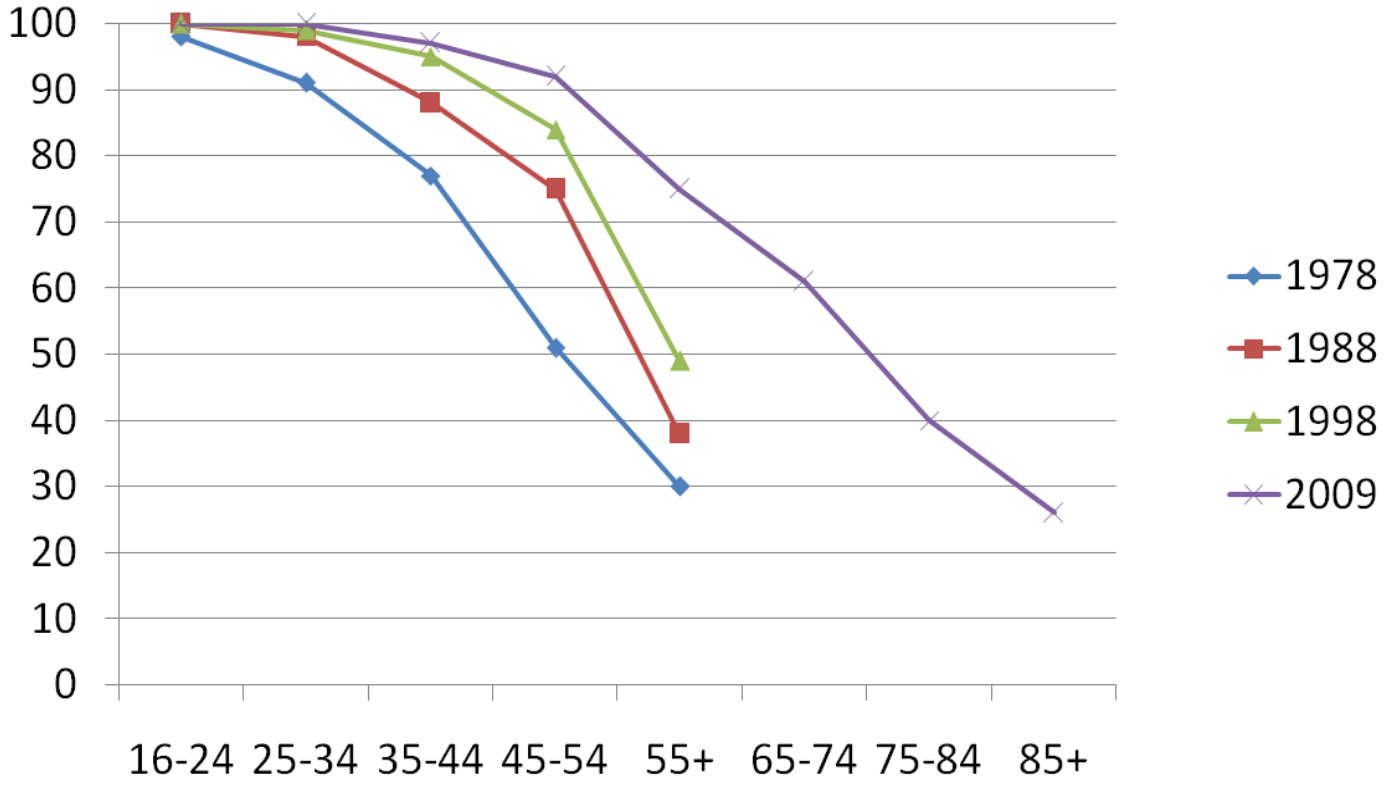
Time-series studies: use

- Compare changes over time
- **Descriptive**: changes in a condition over time in a population
- **Analytical**: relate changes in exposure to changes in outcome
- **Long-term** trends (e.g. lung cancer mortality and smoking rates)
- **Short-term** variation (e.g. daily changes in air pollution and mortality).

Time-series (vs. other ecological) studies

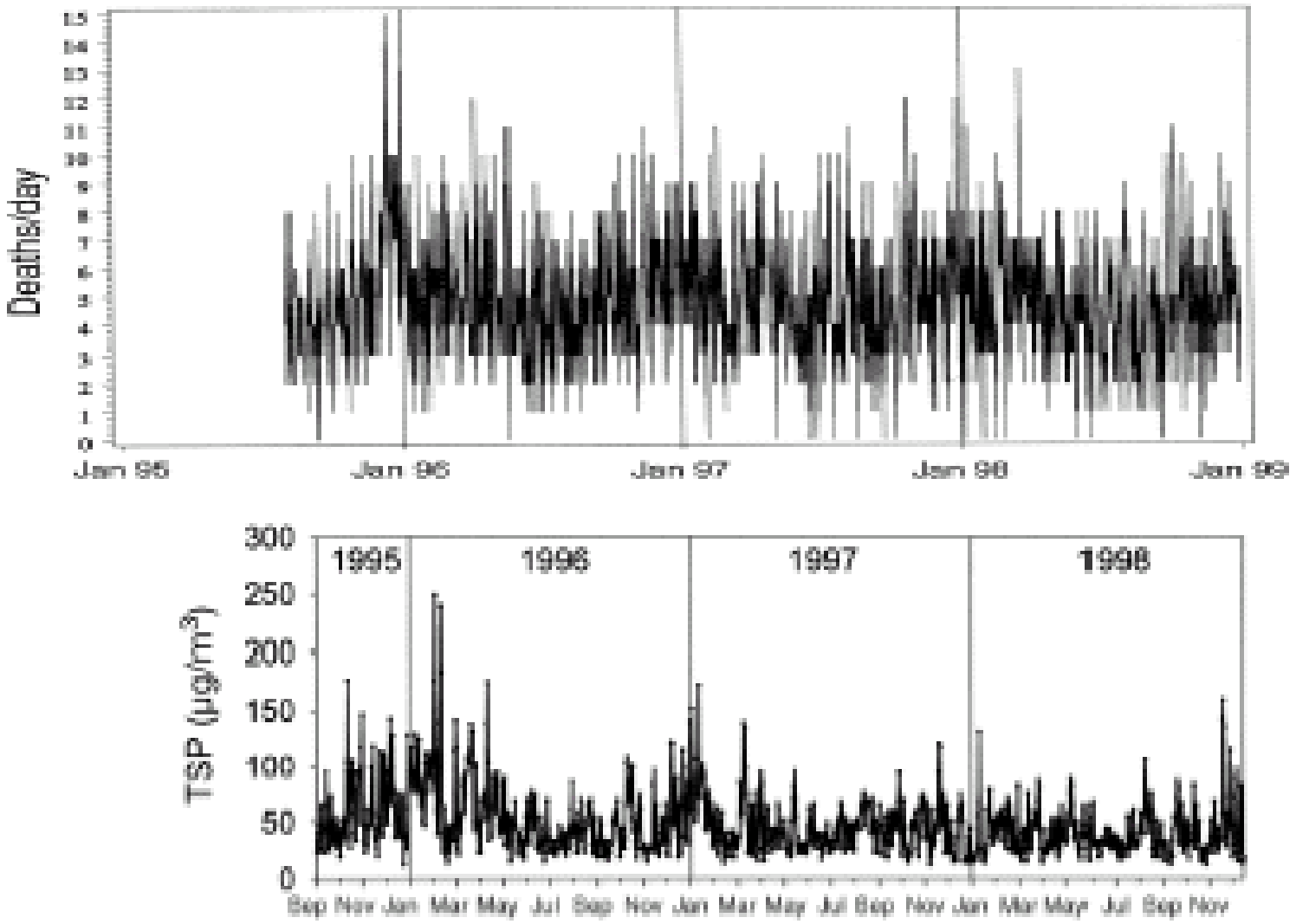
- Advantages
  - help reduce confounding (e.g. it is unlikely that smoking rates would change within a population over a period of several days).
  - Resemble experiment (before and after)
- Disadvantages
  - There can be other factors changing over time - confounding
  - Many exposures influence health with a lag which is often unknown (e.g. pollution and mortality) or very long (e.g. lung cancer and smoking).

Retention of 21+ natural teeth (%): Adult Dental Health Surveys



Fuller E, Steele JG, Watt RG, Nuttall N. Oral health and function – a report from the Adult Dental Health Survey 2009 (www.ic.nhs.uk)

Daily deaths and pollution



From Wichman et al, HEI research report, 2000

Ecological studies: Advantages

- Use existing (often routinely collected) data
- Quick and cheap
- Useful to general hypotheses
- Differences in both exposure and outcome rates may be large, which increases the likelihood to find an association
- Some exposures are difficult to measure in individuals and area-based measures are used instead (e.g. air pollution), and some exposures are inherently ecological (e.g. income inequality)
- Using both ecological and individual level data requires a special type of *multi-level* analyses

Ecological studies: Disadvantages

- Confounding: the groups, which are compared (e.g. countries) usually differ in many other factors than the exposure of interest. It is often impossible to reliably control for confounders.
- There can be systematic differences in measurements of exposures and diseases (e.g. coding of causes of death) between populations.
- Boundaries of different units are sometimes artificial → misleading results.
- Ecological fallacy: ecological studies compare groups but results are extrapolated to individuals.