Bradford Hills criteria (continued)

Is there a valid statistical association?

- Is the association likely to be due to chance?
- Is the association likely to be due to bias?
- Is the association likely to be due to confounding?

Can this valid statistical association be judged as cause & effect?

- Is there a strong association?
- Is there consistency with other studies?
- Is there biological credibility to the hypothesis?
- Is the time sequence compatible?
- Is there evidence of a dose-response relationship?

Interpretation of epidemiological data

- Magnitude of effect
  - Great effect, hardly unknown confounder
  - Is there consistency with other studies?
    - Have others made similar observations?
  - Biologic credibility
  - Is the time sequence sound?
    - Does exposure precede outcome?
  - Is there evidence of a dose-response pattern?
Biological credibility?

- Personal characteristics and skull shape (phrenology)
- Stress and gastric ulcers
- Swimming one hour after eating

’In earlier times we thought that this disease was caused by an evil spirit. Now we know better – it is caused by a garden gnome…’

Interpretation of epidemiological data

- Magnitude of effect
  - Great effect hardly unknown confounder
- Is there consistency with other studies?
  - Have others made similar observations?
- Biologic credibility
  - Is the time sequence sound?
    - Does exposure precede outcome?
  - Is there evidence of a dose-response pattern?

Time sequence

Interpretation of epidemiological data

- Magnitude of effect
  - Great effect hardly unknown confounder
- Is there consistency with other studies?
  - Have others made similar observations?
- Biologic credibility
  - Is the time sequence sound?
    - Does exposure precede outcome?
  - Is there evidence of a dose-response pattern?

Epidemiological way of thought
(Infectious diseases)

- Is there a problem?
- What characterises the problem?
  - When does it occur?
  - Where does it occur?
  - Who’s problem is it?
- Hypothesis (what is the cause of the problem)
  - Is the hypothesis correct?
- Device public health measurements

Descriptive epidemiology

Analytic epidemiology
Diseases can be characterised

- How many?
- Absolute/relative
- Where?
- When?
- Who?
- Gender, age, race, etc..
- Descriptive epidemiology

Disease patterns can be analysed

- Frequency & distribution
- Determinants
- Application

Descriptive and analytic study types

<table>
<thead>
<tr>
<th>Descriptive studies</th>
<th>Analytic studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case reports/series</td>
<td>Case-control studies</td>
</tr>
<tr>
<td>Correlational studies</td>
<td>Cohort studies</td>
</tr>
<tr>
<td>Cross sectional surveys</td>
<td>Randomised/Intervention trials</td>
</tr>
</tbody>
</table>

Case-reports and -series

Carcinoma of the penis and cervix

"... Case 3. – Presented with 5-year history in November, 1969, aged 47. He had massive penile condylomata with squamous carcinomatous change and invaded ingual nodes. Died in 1977. His wife presented with carcinoma of the cervix in 1971 at the age of 43. She had a squamous cell carcinoma and stage III disease. Died 27 months later."

Cartwright and Sinson, 1980; Lancet: 1: 97

Case-reports and -series

"These types of studies in which typically an astute clinician identifies an unusual feature of a disease or a patient's history, may lead to the formulation of a new hypothesis."

"This design has historical importance in epidemiology, as it was often used as an early means to identify the beginning or presence of an epidemic. Investigation of the activities of the affected individuals in the case series can then lead to formulation of a hypothesis."

"While case reports and case series are very useful for hypothesis formulation, they cannot be used to test for the presence of a valid statistical association."
When is too much too much?

- Endemy (sporadic)
- Period/seasonal changes
- Epidemy
- Pandemy

Endemic – sporadic outbreaks

- Sporadic outbreaks that constitute the background frequency (rate) in the population
- Fluctuation (daily/weekly/monthly), but overall not significantly different from background rate
- Constitute the main part of infections in a population

Endemicity

Respiratory tract infection in children in Greenland

Periodic changes

Fig. 3.4 Notifications of measles in England and Wales showing periodic variation (prior to introduction of measles vaccination). Reproduced with permission of the Office of Population Censuses and Surveys ( Crown copyright).

Seasonal variation

The epidemic

“Epidemic... include any disease, infectious or chronic, occurring at a greater frequency than usually expected”

When is that?

- Point source
- Person-to-person (propagated)
When does the observed number exceed the expected?

- 500 cases of pneumonia in Zealand in toddlers January 2001, but 50 cases in June. *Epidemic?*
  - Every winter 500 cases - *RS-virus*

Cases of Kaposi’s sarcoma in S.F.

The epidemic – the special situation

*Source: Centers for Disease Control and Prevention*

Anthrax incidence

*Source: Centers for Disease Control and Prevention*
Surveillance / epidemic

Pandemics

SARS
Spanish flu
Avian flu

Measures of frequency

- **Prevalence (prevalence rate)**
  
  \[
  \text{Number of sick persons at given time} \div \text{Number of persons in the population}
  \]
  
  - **Point prevalence** – prevalence at given time (Christmas eve)
  - **Periodic prevalence** – prevalence in a period (Christmas holiday)

- **Incidence (incidence rate)**

  \[
  \text{Number of new cases of disease in a specific period} \div \text{Sum of time at risk for the population}
  \]

  - **Duration of episode**

The concept of time at risk

Why different measures?

- Prevalence measures the appearance of disease at a specific time in a population
- Measure of burden of disease

- Incidence measures the frequency of disease per unit of time or the risk of getting it
- Measure of risk
Factors affecting prevalence

Prevalence = Incidence * duration of disease

- Longer duration of disease
- Prolongation of life without cure
- Increase in new cases
- Out-migration of cases
- In-migration of susceptibles
- Improved diagnostic facilities
- Shorter duration of disease
- Higher case-fatality rate from disease
- Decrease in new cases
- In-migration of healthy
- Out-migration of cases
- Improved cure rate

What measure to use?

The measure depends on the question!

Incidence and prevalence: practical example

- Respiratory tract infections (RTI) in children in Greenland
- What is the prevalence and incidence of RTI?

Day 1

<table>
<thead>
<tr>
<th>Time at risk</th>
<th>Time of observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
</tr>
</tbody>
</table>

- Prevalence
  - Time with disease by time of observation
  - \(\frac{(15+15+11)}{100} = 41\%\)
  - Measure of disease burden

- Incidence
  - Number of new episodes by time at risk
  - \(3\) per 50 days
  - Measure of risk

Population at risk

- Crucial to the calculation of frequency and rates that the population at risk is defined! Who is at risk?

- Influenza
  - All who have not been infected with the (this year) prevailing serotypes or are unvaccinated

- Cervical cancer
  - Women aged 25-69 years

- Breast cancer
  - All

- Salmonella outbreak in restaurant
  - All who have tasted the food

- Hospital infections
  - Salmonella in the central kitchen
  - Defect bedpan disinfector in ward

Interpretation of prevalence: comparability

- Graphs do not necessarily express different infection patterns
- Cohort effect

Background population

- England 1983: ‘Windscale – the Nuclear Factory’ (Sellafield)
- Statistically significant excess number of cases of childhood leukemia in the village Seascale
- Should the plant be shut down?
The Texas sharpshooter

Risk factors

- A factor associated with an increased risk of disease but not sufficient to cause disease

Possible mechanisms
- Predisposing (sex, age, previous disease)
- ‘Enabling factors’ (low social class, low income, bad nutritional status, bad housing conditions)
- ‘Exposural factors’ (exposure to infectious agent or chemical)
- ‘Enhancing factors’ (hard work (pilots – reduced immune function)

Risk

- Absolute risk
  - Risk of dying if you smoke

- Relative risk
  - Risk of dying if you smoke compared to if you don’t

- Risk difference (attributable risk, risk difference)
  - The extra risk attributable to presence of the factor

Risk example

<table>
<thead>
<tr>
<th>Smokers</th>
<th>Dead</th>
<th>Number of deaths/1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>25,769</td>
<td>133</td>
<td>5.16</td>
</tr>
<tr>
<td>Non-smokers</td>
<td>5,439</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Absolute risk of death smokers? 5.16 per 1000
Relative risk smokers/non-smokers? 5.16/0.55 9.38
Risk difference? 5.15/0.55 4.61 per 1000

Population attributable risk percent

‘Population attributable risk percent’ – theoretical measure of the proportion of disease cases in a population, if the risk factor did not exist

<table>
<thead>
<tr>
<th>Otitis media in Greenland</th>
</tr>
</thead>
</table>
| Ethnicity (2 greenlandic parents, 358/409): | 80%
| Familial disposition (mother ear discharge, 50/430): | 21%
| Childcare |
| – Day care | 6%
| – Childcare center | 76%
| Smokers in household (for children <1 year, 372/465): | 74%

Descriptive and analytic study types

<table>
<thead>
<tr>
<th>Descriptive studies</th>
<th>Analytic studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case reports/series</td>
<td>Case-control studies</td>
</tr>
<tr>
<td>Correlational studies</td>
<td>Cohort studies</td>
</tr>
<tr>
<td>Cross sectional surveys</td>
<td>Randomised/Intervention trials</td>
</tr>
</tbody>
</table>
Ecological studies

In correlational studies, measures that represent characteristics of entire populations are used to describe disease in relation to some factors of interest such as age, calendar time, utilization of health services or consumption of a food, medication or other product. (H&B p 102)

**Correlation**

- $0 < r < 1$
- $-1 < r < 0$
- $r = 0$

**Occurrence**

“Exposure”

Kaposi’s sarcoma

Fig. 1. Kaposi’s sarcoma and non-Hodgkin’s lymphoma incidence among men, per 100,000 people per year, age-standardized to the 1970 U.S. population, shown on a linear and log scale to illustrate both the absolute and relative changes in nine Surveillance, Epidemiology and End Results (SEER) registries and in the San Francisco area registry only, from 1973 through 1998. Years with no cases were set arbitrarily at 0.12 cases in the log scale.

Correlational studies

“The presence of a correlation does not necessarily imply the presence of a valid statistical association. Conversely, lack of a correlation in such studies does not necessarily imply the absence of a valid statistical association.”

(H&B p. 104)

Cross sectional surveys

“A third type of study is the cross-sectional or prevalence survey, in which exposure and disease status are assessed simultaneously among individuals in a well-defined population.”

(H&B p. 108)

Helicobacter pylori seroprevalence in Greenland 1998

- $N=685$
- Average seropositivity 15-85 years: 58.5%

HP risk factors, Greenland 1998

<table>
<thead>
<tr>
<th>No. of children</th>
<th>N (% seropositive)</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25 (16)</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>1</td>
<td>48 (36)</td>
<td>2.9 (0.75 – 11.2)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>44 (27)</td>
<td>(0.38 – 6.07)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>40 (30)</td>
<td>(1.52 – 26.4)</td>
<td></td>
</tr>
<tr>
<td>3+</td>
<td>132 (53)</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>No. of older sibling</td>
<td>74 (36)</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>1-2</td>
<td>97 (32)</td>
<td>1.44 (0.37 – 5.08)</td>
<td></td>
</tr>
<tr>
<td>3+</td>
<td>22 (64)</td>
<td>8.58 (1.45 – 50.7)</td>
<td></td>
</tr>
<tr>
<td>Distance to nearest older sibling</td>
<td>0-1 years</td>
<td>32 (53)</td>
<td>1</td>
</tr>
<tr>
<td>2 years</td>
<td>22 (36)</td>
<td>0.57 (0.1 – 3.37)</td>
<td></td>
</tr>
<tr>
<td>3-4 years</td>
<td>31 (42)</td>
<td>1.31 (0.3 – 5.66)</td>
<td></td>
</tr>
<tr>
<td>5+ years</td>
<td>34 (21)</td>
<td>0.17 (0.03 – 0.9)</td>
<td></td>
</tr>
</tbody>
</table>
Cross sectional surveys

- Studies, in which outcome and exposure are determined simultaneously
- The observed outcomes are prevalent
- Data on risk factor associations will accordingly represent both survival and etiology
- It cannot be ruled out that the exposure under observation has changed after and maybe because of the outcome
- It cannot always be determined which came first, the exposure or the outcome

Descriptive vs. analytic epidemiology

**Descriptive epidemiology**

**Advantages**
- Cheap & quick
- May provide important overview

**Disadvantages**
- No information on the individual
- May involve bias
- Results may be ambiguous
- Can not test (causal) hypotheses

**Analytic epidemiology**

**Advantages**
- Information on the individual
- Control for confounding
- Results less ambiguous
- Can test (causal) hypotheses

**Disadvantages**
- Expensive
- Laborious
- May involve bias

Analytic study types

- **Cohort studies**
  - Information on exposure and outcome in the whole study population
  - Frequent outcomes
- **Case-control studies**
  - Only information about sample from the population
  - More rare disease
- **Randomised controlled studies**

Two different situations

Exposure

- **Cohort**
- **Case-control**

Cohort studies

- **Cohort**: Cohors (latin): 10th of legion
- Prospektively (!)
- Starting point population of healthy persons

**Cohort studies measure**

- Risk of disease among exposed compared with risk of disease among unexposed
- Relative risk
- Absolute risk can be determined for both groups
- When outcome is frequent and cohort study practically feasible
Case-control studies

- Starting point patients with disease
- Compared with control group (retrospective (!))
- Frequency of risk factor (‘cause’) among patients (cases) compared with frequency among controls

Time

<table>
<thead>
<tr>
<th>Exposed</th>
<th>Non-exposed</th>
<th>Cases (with disease)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposed</th>
<th>Non-exposed</th>
<th>Controls (without disease)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Population

Questioning

Case-control studies measure

- Degree of exposure among the sick compared with the degree of exposure among controls
- Odds ratio (no unit)
- Rates (~risks) cannot be computed as only samples are drawn from the population
- When outcome is less frequent and there are practical constraints to do a cohort study

Randomised controlled study

- Group of patients/persons where half are allocated by lottery to treatment, the other to placebo
- What is the big advantage?
- Confounder control!
  - How, and why is that smart?

Take home messages

- Epidemiological methods can be used to identify and characterise diseases, their causes and natural history
- Descriptive methods generate hypotheses, but cannot prove them
- Analytic methods (cohort, case-control and randomised clinical trials) test hypotheses
- Each method has advantages and disadvantages that determines the actual use

Applications of epidemiology

Epidemiological methods can be used to
- Identify (new) diseases
- Characterise the natural history of diseases
- Characterise disease occurrence in populations
- Identify causes of diseases
- Evaluate the efficacy and effectiveness of interventions

WHO issues a global alert about cases of atypical pneumonia

“In Viet Nam the outbreak began with a single initial case who was hospitalized for treatment of severe, acute respiratory syndrome of unknown origin. He felt unwell during his journey and fell ill shortly after arrival in Hanoi from Shanghai and Hong Kong SAR, China. Following his admission to the hospital, approximately 20 hospital staff became sick with similar symptoms.”

WHO issues a global alert about cases of atypical pneumonia

"The signs and symptoms of the disease...include initial flu-like illness (rapid onset of high fever followed by muscle aches, headache and sore throat)...most common symptoms. Early laboratory findings may include thrombocytopenia (low platelet count) and leucopenia (low white blood cell count)...some, but not all...followed by bilateral pneumonia, in some cases progressing to acute respiratory distress requiring assisted breathing on a respirator. Some patients are recovering but some patients remain critically ill."


Applications

Epidemiological methods can be used to
• Identify (new) diseases
• Characterise the natural history of diseases
• Characterise disease occurrence in populations
• Identify causes of diseases
• Evaluate the efficacy and effectiveness of interventions

Outcomes and Prognostic Factors in 257 Patients with Severe Acute Respiratory Syndrome in Hong Kong

Chui KW et al., Ann Intern Med. 2003 Nov 4;139(9):715-23

Applications

Epidemiological methods can be used to
• Identify (new) diseases
• Characterise the natural history of diseases
• Characterise disease occurrence in populations
• Identify causes of diseases
• Evaluate the efficacy and effectiveness of interventions

Applications

Epidemiological methods can be used to

• Identify (new) diseases
• Characterise the natural history of diseases
• Characterise disease occurrence in populations
• Identify causes of diseases
• Evaluate the efficacy and effectiveness of interventions

Identification of severe acute respiratory syndrome in Canada

SARS was first identified in Canada in early March 2003. We collected epidemiologic, clinical, and diagnostic data from each of the first 10 cases prospectively as they were identified. Specimens from all cases were sent to local, provincial, national, and international laboratories for studies to identify an etiologic agent.

Patients died, and five have had clinical improvement. The results of laboratory investigations were negative or not clinically significant except for the amplification of human metapneumovirus from respiratory specimens from five of six patients and from one asymptomatic contact of a patient with SARS. A novel coronavirus was isolated and amplified from respiratory specimens from five of six patients. In four cases both pathogens were isolated.


Applications

And the next epidemiology lessons…

• Much more about analytic studies!