Epidemiologic Study Designs

Visanou HANSANA
MD; MCTM

GFMER - WHO - UNFPA - LAO PDR
Training Course in Reproductive Health Research
Vientiane, 13 October 2009
Epidemiologic Study Designs

Experimental (RCTs)

Observational

Analytical

Descriptive

Case-Control

Cohort

+ cross-sectional & ecologic
Descriptive studies
Examine patterns of disease

Analytical studies
Studies of suspected causes of diseases

Experimental studies
Compare treatment modalities
Epidemiologic Study Designs

Grimes & Schulz, 2002
Hierarchy of Epidemiologic Study Design

Generate hypotheses

Establish causality

Case reports
Case series
Ecologic studies
Cross-sectional studies
Case-control studies
Cohort studies
Randomized controlled trials

Tower & Spector, 2007 (www)
Observational Studies
(no control over the circumstances)

- **Descriptive**: Most basic demographic studies

- **Analytical**: Comparative studies testing an hypothesis
  * cross-sectional
    (a snapshot; no idea on cause-and-effect relationship)
  * cohort
    (prospective; cause-and-effect relationship can be inferred)
  * case-control
    (retrospective; cause-and-effect relationship can be inferred)
Figure 2: Schematic diagram showing temporal direction of three study designs
Analytical Studies
(comparative studies testing an hypothesis)

* **cohort** (prospective)
  
  *Begins with an exposure (smokers and non-smokers)*

* **case-control** (retrospective - trohoc)
  
  *Begins with outcome (cancer cases and healthy controls)*
<table>
<thead>
<tr>
<th>Population</th>
<th>People without disease</th>
<th>Exposed</th>
<th>People without disease</th>
<th>Disease</th>
<th>No disease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Not exposed</td>
<td></td>
<td>Disease</td>
<td>No disease</td>
</tr>
</tbody>
</table>
Examples of Cohort Studies

* Framingham Heart Study  [www](www)
* NHANES Studies  [www](www)
* MACS  [www](www)
* Physicians' Health Study  [www](www)
* Nurses' Health Study  [www](www)
* ALSPAC  [www](www)
Advantages of Cohort Studies

- Can establish population-based incidence
- Accurate relative risk (risk ratio) estimation
- Can examine rare exposures (asbestos > lung cancer)
- Temporal relationship can be inferred (prospective design)
- Time-to-event analysis is possible
- Can be used where randomization is not possible
- Magnitude of a risk factor’s effect can be quantified
- Selection and information biases are decreased
- Multiple outcomes can be studied (smoking > lung cancer, COPD, larynx cancer)
Disadvantages of Cohort Studies

- Lengthy and expensive
- May require very large samples
- Not suitable for rare diseases
- Not suitable for diseases with long-latency
- Unexpected environmental changes may influence the association
- Nonresponse, migration and loss-to-follow-up biases
- Sampling, ascertainment and observer biases are still possible
Presentation of cohort data:
Population at risk

Does HIV infection increase risk of developing TB among a population of drug users?

<table>
<thead>
<tr>
<th>Population (follow up 2 years)</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV +</td>
<td>215</td>
</tr>
<tr>
<td>HIV -</td>
<td>289</td>
</tr>
</tbody>
</table>

Source: Selwyn et al., New York, 1989
Does HIV infection increase risk of developing TB among drug users?

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Population (f/u 2 years)</th>
<th>Cases</th>
<th>Incidence (%)</th>
<th>Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV +</td>
<td>215</td>
<td>8</td>
<td>3.7</td>
<td>11</td>
</tr>
<tr>
<td>HIV -</td>
<td>298</td>
<td>1</td>
<td>0.3</td>
<td></td>
</tr>
</tbody>
</table>

EPIET (www)
Presentation of cohort data: Person-years at risk

Tobacco smoking and lung cancer, England & Wales, 1951

<table>
<thead>
<tr>
<th>Person-years</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke</td>
<td>102,600</td>
</tr>
<tr>
<td>Do not smoke</td>
<td>42,800</td>
</tr>
</tbody>
</table>

Source: Doll & Hill
## Presentation of data: Various exposure levels

<table>
<thead>
<tr>
<th>Daily number of cigarettes smoked</th>
<th>Person-years at risk</th>
<th>Lung cancer cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 25</td>
<td>25,100</td>
<td>57</td>
</tr>
<tr>
<td>15 - 24</td>
<td>38,900</td>
<td>54</td>
</tr>
<tr>
<td>1 - 14</td>
<td>38,600</td>
<td>22</td>
</tr>
<tr>
<td>none</td>
<td>42,800</td>
<td>3</td>
</tr>
</tbody>
</table>
## Cohort study: Tobacco smoking and lung cancer, England & Wales, 1951

<table>
<thead>
<tr>
<th>Cigarettes smoked/d</th>
<th>Person-years at risk</th>
<th>Cases</th>
<th>Rate per 1000 p-y</th>
<th>Rate ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 25</td>
<td>25,100</td>
<td>57</td>
<td>2.27</td>
<td>32.4</td>
</tr>
<tr>
<td>15 - 24</td>
<td>38,900</td>
<td>54</td>
<td>1.39</td>
<td>19.8</td>
</tr>
<tr>
<td>1 - 14</td>
<td>38,600</td>
<td>22</td>
<td>0.57</td>
<td>8.1</td>
</tr>
<tr>
<td>none</td>
<td>42,800</td>
<td>3</td>
<td>0.07</td>
<td>Ref.</td>
</tr>
</tbody>
</table>

Source: Doll & Hill
Prospective cohort study

Exposure

Study starts

Disease occurrence

EPIET (www)

Prospective cohort study

Study starts

Exposure

Disease occurrence

EPIET (www)
Retrospective cohort studies

Exposure → Disease occurrence → Study starts

(time)
Figure 2: Schematic diagram of concurrent, retrospective, and ambidirectional cohort studies
Case-Control Studies

Exposed
Not exposed

Exposed
Not exposed

Cases

Population

Exposed

Controls
Case-Control Studies

Schulz & Grimes, 2002 (www) (PDF)
Advantages of Case-Control Studies

- Cheap, easy and quick studies
- Multiple exposures can be examined
- Rare diseases and diseases with long latency can be studied
- Suitable when randomization is unethical (alcohol and pregnancy outcome)
Disadvantages of Case-Control Studies

- Case and control selection troublesome
- Subject to bias (selection, recall, misclassification)
- Direct incidence estimation is not possible
- Temporal relationship is not clear
- Multiple outcomes cannot be studied
- If the incidence of exposure is high, it is difficult to show the difference between cases and controls
- Not easy to estimate attributable fraction
- Reverse causation is a problem in interpretation - especially in molecular epidemiology studies
# Case-Control Studies: Potential Bias

## Panel 2: Introduction of bias through poor choice of controls

<table>
<thead>
<tr>
<th>Cases</th>
<th>Control selection</th>
<th>Non-representativeness</th>
<th>Selection bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal cancer patients admitted to hospital</td>
<td>Patients admitted to hospital with arthritis</td>
<td>Controls probably have high degrees of exposure to NSAIDs</td>
<td>Would spuriously reduce the estimate of effect (odds ratio)</td>
</tr>
<tr>
<td>Colorectal cancer patients admitted to hospital</td>
<td>Patients admitted to hospital with peptic ulcers</td>
<td>Controls probably have low degrees of exposure to NSAIDs</td>
<td>Would spuriously increase the estimate of effect (odds ratio)</td>
</tr>
</tbody>
</table>

NSAIDs = non-steroidal anti-inflammatory drugs.
Epidemiologic Association / Impact Measures

(Absolute Risk) (AR)
Relative Risk (Risk Ratio) (RR)
Odds Ratio (OR)

Measures of test accuracy:
Sensitivity, specificity, positive and negative predictive value (PPV, NPV)
### Odds Ratio: 3.6
95% CI = 1.3 to 10.4

---

**Association Studies**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Type 1</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA DR4</td>
<td>17</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>NON-HLA DR4</td>
<td>20</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>37</strong></td>
<td><strong>37</strong></td>
<td><strong>74</strong></td>
</tr>
</tbody>
</table>

\[ X^2 = 5.377 \]

\[ p < 0.025 \]

**ROCHE Genetic Education** (www)
OR = \frac{ad}{bc} = \frac{17 \times 30}{20 \times 7} = 3.6

RR = \frac{a}{(a+c)} \div \frac{b}{(b+d)} = \frac{17/24}{20/50} = 1.8

EBM toolbox (www)
EpiMax Table Calculator (www)
Epidemiologic Study Designs

Figure 3: Algorithm for distinguishing rates, proportions, and ratios
Sources of Error in Epidemiologic Studies

Random error
Bias
Confounding
Effect Modification
Reverse Causation
Sources of Error in Epidemiologic Studies

Random error

*Large sample size, replication*

Bias

*Be careful*

Confounding

Effect Modification

Reverse Causation
Confounding can be controlled by:

- **Randomization**: assures equal distribution of confounders between study and control groups

- **Restriction**: subjects are restricted by the levels of a known confounder

- **Matching**: potential confounding factors are kept equal between the study groups

- **Stratification** for various levels of potential confounders

- **Multivariable analysis** (does not control for effect modification)
Effect modification can be assessed by:
- **Stratification** for various levels of potential confounders
- **Multivariable analysis** (by assessing interaction)

Reverse causation can be assessed by:
- **Mendelian Randomization**
Thank you