Overview of Study Designs

Shah Islam, PStat, MPH
Department of Biostatistics
What are most studies trying to determine

Exposure

Disease
Or
Other Outcomes

Association
Classification of Study Designs

Observational
- Descriptive
  - Case reports and series
  - Ecological studies
  - Cross sectional studies
  - Case-Control
  - Cohort
- Analytic
- Randomized Controlled Trials
- Others (e.g. before-after studies and non-randomized trials)

Experimental
Observational Studies

- Considered “natural” experiments because investigator lets nature take its course

- Take advantage of fact that people expose themselves to noxious or healthy substances through personal habits, choice of occupation, place of residence, and so on.

- Provide information on exposures that occur in natural settings and are not limited to preventions and treatments
Descriptive Studies

- Describe patterns of disease occurrence in relation to person, place and time

- Provide valuable information to enable health care providers to allocate resources efficiently and to plan effective prevention or education programs

- Have often provided the first important clues about possible determinants of disease

- Are primarily useful for formulation of hypothesis
Case Reports

- Most basic type of descriptive study of individuals

- Consists of careful, detailed report by one or more clinicians of profile of single patient

- For example, in 1961 a case-report was published of 40-year-old premenopausal women who developed a pulmonary embolism 5 weeks after beginning to use an oral contraceptive
Limitations of Case Reports

- Cannot be used to test for presence of a valid statistical association
- Based on the experience of only one person
- The presence of any risk factor, however suggestive, may simply be coincidental
Case Series

- Collections of individual case reports, usually occurring in a fairly short period of time

- Often used as an early means to identify the beginning of presences of an epidemic

- For example, five young, previously healthy homosexual men were diagnosed as having Pneumocystis carinii pneumonia at three LA hospitals during a 6-month period in 1980-1981
Limitation of Case Series

- Cannot be used to test for presence of a valid statistical association

- Interpretability of information is severely limited by lack of an appropriate comparison group

- This lack can either obscure a relationship or suggest an association where none actually exists
Correlational Studies

SOME KEY POINTS

- Simultaneously assess exposure and disease
- Establishes the **Strength** and **direction** of association between exposure and disease
- Used to generate hypothesis **NOT** to test causality

Can be done on 2 levels...

1. Aggregate data – Ecologic Studies
2. Individual data – Cross-sectional Studies
Ecologic Studies

- The measures are presented as summary statistics such as means, median, or mode.

- The design tends to be used to generate hypotheses about the relationship between exposures and outcomes.

- A simple correlation coefficient is used as the measure of association.

- **Example**: A study of mortality from lung disease in different cities that are known to have different levels of air pollution would comprise an ecologic study. The unit of analysis is a city.
Ecologic Fallacy

- **definition** – the belief that a correlation between an exposure and an outcome at the population level occurs at the individual level, when it does not.

- **implication** – group level data cannot be used for hypothesis-testing unless it is a summary of individual measures.
Correlation between fat intake and breast cancer by country

From Gordis, *Epidemiology*, Figure 13.4
Strengths of Ecologic Studies

- can be done quickly and inexpensively, often using already available information

- surveillance data can permit comparisons of disease rates in different geographic regions

- analysis and presentation are relatively simple and easy to understand

- ability to achieve a wider range of exposure levels than could be expected from a typical individual-level study
Limitations of Ecologic Studies

- Lack of individual-level information leads to “ecological fallacy”

- Correlational data represent average exposure levels rather than actual individual values

- Investigators’ inability to detect subtle or complicated relationships (such as J-shaped or other curvilinear relationships)

- Lack of information on potential confounding factors
Cross-sectional Studies

- exposure and disease are assessed simultaneously in each INDIVIDUAL at a given point or SNAPSHOT in time...

ONE SLICE IN TIME
Cross-Sectional Study - I

Begin with:

- Defined Population

Gather Data on Exposure and Disease

Four Groups Are Possible:

- Exposed; Have Disease
- Exposed; Do not Have Disease
- Not Exposed; Have Disease
- Not Exposed; Do not Have Disease
Cross-Sectional Study - III

Prevalence of disease in exposed compared to non-exposed

\[
\frac{a}{a+b} \quad \text{vs.} \quad \frac{c}{c+d}
\]

Prevalence of exposure in diseased and non-diseased

\[
\frac{a}{a+c} \quad \text{vs.} \quad \frac{b}{b+d}
\]
Strengths of Cross-sectional Studies

- Quick, requiring only a “one-time” examination/interview
- Less expensive than other designs
- Helpful in program planning and determining types of health services needed
- Useful in determining associations between variables of interest, thereby generating hypotheses
Limitations of Cross-sectional Studies

- Does not separate cause-effect relationships in associations established; temporal relationships unclear
- Deals only with survivors
- Not useful when rare health conditions are being considered
- Does not identify risk or future likelihood of occurrence disease from given characteristic
- Has limited usefulness for explosive epidemics or acute, short duration illnesses such as measles or upper respiratory infections
Analytic Studies

- Comparison is explicit, since investigator Assembles groups of individuals for specific purpose of systematic determination of whether or not Risk of disease differs for exposed versus unexposed

- Use of an appropriate comparison group allows testing of epidemiologic hypotheses in analytic study designs
Case-Control Studies

Some key points

❖ Most frequently used study design

❖ Participants selected on the basis of whether or not they are DISEASED (note that in a cohort study participants are selected based on exposure status)

❖ Those who are diseased are called CASES

❖ Those who are not diseased are called CONTROLS
Design of a Case-control Study

Exposed → Disease → "CASES"
Exposed → No Disease → "CONTROLS"
Schematic of Case-control Study

1. Take histories
2. Compare histories
3. Draw conclusions

Group of interest (e.g. cancer patients)

Comparison group (e.g. non-patients)
Participants are selected on the basis of disease, exposures for all participants are obtained retrospectively.

Example: Lung cancer cases and non-cancerous controls recall past exposure to cigarette smoke.
Measures of Association

• Odds Ratio (OR)
Assume a study of 10 cases and 10 *unmatched* controls, with the following findings:

<table>
<thead>
<tr>
<th>CASES</th>
<th>CONTROLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>N</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>E</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>E</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>E</td>
<td>N</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>E</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
Thus, 6 of the 10 cases were exposed and 3 of the 10 controls were exposed.

Setting up these data in a $2 \times 2$ table, we obtain:

<table>
<thead>
<tr>
<th></th>
<th>CASES</th>
<th>CONTROLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CASES</th>
<th>CONTROLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Cases</td>
<td>Controls</td>
</tr>
<tr>
<td>----------------</td>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td>Exposed</td>
<td>6 (a)</td>
<td>(b) 3</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>4 (c)</td>
<td>(d) 7</td>
</tr>
</tbody>
</table>

10 10

The odds ratio in this *unmatched* study

= the ratio of the cross products:

\[
\frac{ad}{bc} = \frac{6 \times 7}{4 \times 3} = \frac{42}{12} = 3.5
\]

In case of matched study OR = b/c
Strengths of Case-control Studies

- Good for studying rare diseases
- Can use smaller sample sizes
- Cost/time effective when using previously collected (RETROSPECTIVE) exposures
- Tend to support causal hypotheses by establishing associations
Limitations of Case-control Studies

- Subject to bias (selection and recall)
- Can’t calculate incidence
- Selecting appropriate controls can be challenging
- Data from clinical records may be inadequate or incomplete
Cohort Studies

- definition: A study in which two or more groups of people that are free of disease and that differ according to the extent of exposure (e.g. exposed and unexposed) are compared with respect to disease incidence
Types of Cohort Studies

- **Prospective (concurrent):**
  - Begin with a group of exposed and unexposed and follow them for years

- **Retrospective (non-concurrent):**
  - Look back into the past to find a group of exposed and unexposed and ‘follow’ them to the present
Cohort Study Designs

Retrospective

Select groups in 2012

Trace

1970
2012
2042
Cohort Study Designs

Prospective

Select groups in 2012

1970  2012  2042

Follow
Strength of Cohort Studies

- Maintain temporal sequence between exposure & outcome
- Assess the various effects of a particular exposure
- Study new or rare exposures
- Calculate measures of risk
- Avoid bias in the exposure measurement
- Provides a complete picture of experience following exposure
Limitations of Cohort Studies

- Likely to be large and expensive
- Inefficient for studying rare diseases
- Potentially long duration of follow-up for some outcomes
- Loss to follow up of subjects
- Must anticipate secular trends in technology, behaviors, etc...
- Difficulties in measuring confounding variables
- Exposures can change through study
Measures of association

<table>
<thead>
<tr>
<th></th>
<th>Diseased</th>
<th>Not Diseased</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Relative Risk (RR) = \( \frac{a/(a + b)}{c/(c + d)} \)

Odds Ratio (OR) = \( \frac{ad}{bc} \)
Experimental Studies

- Studies preventions and treatments for diseases

- Investigator actively manipulates which groups receive the intervention under study
Randomized Controlled Trials (RCT)

- Are generally regarded as the most scientifically rigorous method of hypothesis testing available.

- It is a gold standard of clinical/epidemiological studies
Objectives of Randomized Trials

- Evaluate therapeutic and preventive aspects of medical practice
- Evaluate new approaches to health care delivery
- Evaluate impact of health education health behavior
Reference Population

↓

Study Population

RANDOMIZED

↓

New treatment

Current treatment
Study Population

RANDOMIZATION

New Treatment
- Improved
- Not Improved

Current Treatment
- Improved
- Not Improved
Strength of Clinical Trials

- Considered as providing most reliable evidence from epidemiologic research

- Randomization offers ability to control both known and unknown influences (confounders)

- When well-designed and conducted can provide most direct evidence on which to judge whether an exposure causes or prevents a disease
Limitations of Clinical Trials

- Are time-consuming and usually costly
- Can study only interventions or exposures that are controlled by investigator
- May have problems related to therapy changes and dropouts
- May be limited in generalizability
- Are often unethical to perform at all
Questions?