Lesson 1: Introduction

Lesson slides for this course have been compiled from current and previous instructors of PubH 6414, University of Minnesota, School of Public Health: Ann Brearley PhD MS, Cynthia Davey MS, Susan Telke MS, Judy Bebchuck PhD, Melanie Wall PhD, Chap T. Le PhD
Topics Outline

• The Cycle of Research

• Understanding the Design of a Research Study
  – Sampling Methods
  – Study Designs
  – Measurement Scales
2010 Data: Have you ever been told by a doctor that you have Diabetes?

Responses: Yes, Yes – pregnancy related, No, No – pre-diabetes or borderline.

http://www.cdc.gov/brfss/index.htm
Cycle of Research
Topics Outline

• The Cycle of Research

• Understanding the Design of a Research Study
  – Sampling Methods
  – Study Designs
  – Measurement Scales
Sampling Methods

Study results → Population conclusions

• Requires a representative sample

• Random samples, or ‘probability samples’, tend to be representative
Random Sampling Methods

- Simple random sampling
- Systematic random sampling
- Stratified random sampling
- Cluster random sampling
Simple Random Sampling

A sample of size $n$ is selected by a procedure that gives every sample of size $n$ the same probability of being selected.

– Representative
– Eliminates selection bias
Selection Bias

Selection Bias: “A systematic tendency to favor the inclusion in a sample of selected subjects with particular characteristics while excluding those with other characteristics.” [Source: Pocket Dictionary of Statistics]

– The sample is not representative of the population
– The study results ought not be generalized to the population
How to obtain a Simple Random Sample
Replacement

- Sampling with replacement
- Sampling without replacement

- Which do you suppose is most common?
- Why is that not a problem?
Systematic Random Sampling

A sample of size of \( n \) is selected from a population of size \( N \), by selecting every \( k \)th subject, where \( k = \frac{N}{n} \)

- Why might this type of sampling lead to bias?
Stratified Random Sampling

The population is divided into groups with similar characteristics. A simple random sample is selected from each group.

• What do you suppose the motivation for this type of sampling is?

• What might some common stratification variables be?
Cluster sampling

The population is divided into clusters and the clusters into sub-clusters. Clusters are selected randomly, and within each cluster, sub-clusters are selected randomly.

• When might this type of sampling be used?
Non-Random Sampling

• Quota samples
• Convenience samples

• Problems with non-random sampling
  – Potential for bias
  – Unknown precision
Consider the Sampling Method

• Is the study sample representative of the population of interest?

• Was a random sampling method used?

After giving his friends the free shirts, Ralph lived a long and happy life.
Topics Outline

• The Cycle of Research

• Understanding the Design of a Research Study
  – Sampling Methods
  – Study Designs
  – Measurement Scales
Does the investigator assign exposure?

No

Observational Study

- No Comparison Group
  - Case Series

- Comparison Group
  - Cross-Sectional
  - Case Control
  - Cohort

Yes

Experimental Study

- No Comparison Group

- Comparison Group
  - Not Randomized
    - Uncontrolled Intervention Trial
  - Randomized
    - Randomized Controlled Clinical Trial

Case Series Study

• A series of subjects are observed and the outcomes or other subject characteristics are described
  – Small number of subjects

• Observations from a case series may lead to research questions for more rigorously designed studies
Dr. Harold Jaffe

Centers for Disease Control

Case Definition *(Jaffe et al.)*

**Eligibility Criteria:**
1. Biopsy proven Kaposi sarcoma (KS) or
2. Biopsy, histology, or culture proven infection moderately predictive of cellular immune deficiency
   a. Protozoal and helminthic infections, e.g., *Pneumocystis carinii* pneumonia (now *Pneumocystis jiroveci* and recognized to be a fungal disease)
   b. Fungal infections, e.g., esophageal candidiasis
   c. Bacterial infections, e.g., atypical mycobacterial disease
   d. Viral infections, e.g., cytomegalovirus disease

**Exclusion Criteria:**
1. Previous immunosuppressive therapy
2. Illness associated with immunosuppression
3. Persons with KS over 60 years of age
4. Persons under 10 years of age
Surveillance (*Jaffe et al.*)

1. Review of selected cancer tumor registries

2. Contact with selected physicians in 18 communities

3. Review of request for pentamidine isethionate by CDC Parasitic Diseases Drug Service

4. Reports from individual physicians and state health departments
Results (Jaffe et al.)

- 727 cases were homosexual or bisexual males; of remaining cases 155 were IV drug users.

- 75% of homosexual men were from New York or California.

- 48% of cases were 30 - 39 years of age.

- There was an excess of KS among homosexual / bisexual men.

- Distribution of cases by risk group changed over time.
Case Series Summary

- A descriptive study, i.e., not designed to address a specific hypothesis
  - Clear, reproducible case definition with a focus on person, place and time
  - Thorough surveillance
  - Description of case characteristics over time
  - Identification of future high-risk groups

- Addresses 5 “W” questions: who, what, why, when, and where (Grimes)
Case Series Summary

Weaknesses

• No concurrent control
• Usually inadequate information about persons not developing the disease
• Cases may not be representative; persons with disease are survivors of unknown population

Final Thought

• Often the 1st approach taken and the results are used to generate specific hypotheses
Cross-Sectional Study

- Data are collected from a sample of subjects at one point in time
  - Often a large number of subjects

- Used to explore relationships between variables at a point in time
  - Outcome variables
  - Explanatory variables
  - Demographic characteristics
2010 Data: Have you ever been told by a doctor that you have Diabetes?

Responses: Yes, Yes – pregnancy related, No, No – pre-diabetes or borderline.
## Cross-Sectional Studies

<table>
<thead>
<tr>
<th></th>
<th>Disease Yes</th>
<th>Disease No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure Yes</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>Exposure No</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td>n</td>
</tr>
</tbody>
</table>
Cross-Sectional Study: Examples

• Prevalence studies
  – Point in time
  – Repeated points in time

• Association studies

• Diagnostic or screening test studies
Cross-sectional Study: Pros and Cons

• Pros:
  – Relatively quick and inexpensive

• Cons:
  – Not useful for identifying ‘cause-effect’ relationships
  – Potential for bias
    • Volunteer bias
    • Non-response bias
Case-Control Study

• Subjects are selected based on the outcome
  – ‘Cases’ – have the disease
  – ‘Controls’ – do not have the disease

• Past exposure to suspected risk factor(s) is ascertained
  – Interviews with subjects or family members
  – Medical records
Case-Control Study: Example

• **Cases** and Controls were identified
  – 782 cases with various types of brain tumors
  – 799 controls with nonmalignant conditions

• **Exposure** to cell phone use was evaluated by subject interview

• Study conclusion: “There was no indication of higher brain tumor risk among persons who had used hand-held cellular phones compared to those who had not used them.”

## Case-Control Studies

<table>
<thead>
<tr>
<th></th>
<th>Disease Yes</th>
<th>Disease No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exposure Yes</strong></td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td><strong>Exposure No</strong></td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>a+c</td>
<td>b+d</td>
<td>n</td>
</tr>
</tbody>
</table>
Case-Control Study: Pros and Cons

• Pros:
  – Can obtain an adequate number of cases for studies of rare diseases
  – Economical and quick compared to prospective studies

• Cons:
  – Not useful for identifying ‘cause-effect’ relationships
  – Potential for bias
    • misclassification bias
    • recall bias
  – Cannot be used to estimate prevalence
  – Selection of appropriate controls can be difficult
Cohort Study

• A group of participants (a cohort) is followed over time
• Participants are classified based on their exposure
• The study follows the cohort over time to investigate the association between exposures and outcomes
Cohort Study: Types

• Prospective
  – Exposure has already occurred but outcome has not

• Retrospective
  – Exposure and outcome have both already occurred
Cohort Study: Example

• National Children’s Study (NIH)
  – 100,000 children
  – 105 study locations
  – Followed for 21 years
  – Will examine the effects of environmental influences on child health and development.

• [http://www.nationalchildrensstudy.gov](http://www.nationalchildrensstudy.gov)
# Cohort Studies

<table>
<thead>
<tr>
<th></th>
<th>Disease Yes</th>
<th>Disease No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exposure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>a+c</td>
<td>b+d</td>
<td>n</td>
</tr>
</tbody>
</table>
Cohort Study: Pros and Cons

• Pros:
  – Provide evidence for possible ‘cause-effect’ relationships

• Cons:
  – Time-consuming and expensive
  – Potential for high drop-out or loss-to-follow-up rates
  – Not suitable for very rare outcomes
  – Retrospective cohort design requires complete and accurate medical records
Experimental Studies

• Controlled
  – Involves control subjects who receive
    • No intervention
    • Placebo
    • Standard treatment (if any)
    • Intervention at a later time

• Uncontrolled
Controlled Clinical Trials: Study Designs

- Concurrent
- Self-controlled
- Cross-over
Controlled Clinical Trials: Blinding

- Single-blind
- Double-blind
- Triple-blind
Controlled Clinical Trials: Random Assignment

• Participants are randomly assigned to the intervention group or the control group
  – Randomized controlled clinical trial or RCCT

• Observational studies may use random sampling

• Randomized controlled clinical trials use random assignment
RCCT: Example

• MRFIT: Multiple Risk Factor Intervention Trial
  – 12,866 men with risk factors for CHD
  – Randomized to special intervention vs. usual care
  – Endpoint: death from CHD
  – 6-8 years of follow-up

• Study results
  – Risk factor levels declined slightly more in the special intervention group
  – Mortality was not significantly different

RCCTs: Pros and Cons

• Pros:
  – Provide the strongest evidence for cause-effect relationships

• Cons:
  – Time-consuming and expensive
  – Potential for bias
    • Procedure Bias
    • Recall bias
    • Compliance bias
Consider the Study Design

• For an observational study:
  – What study design was used?
  – Was random sampling used?
  – What are the potential sources of bias?

• For an experimental study:
  – What study design was used? Controlled? Blinded?
  – What are the characteristics of the study sample?
  – Was random assignment used?
  – What are the potential sources of bias?
Hierarchy of Evidence for Making Treatment Guidelines

Coherence of evidence from multiple sources

Systematic review of well-designed, large randomized trials

Strong evidence from one large randomized trial

Systematic review of small trials (e.g., surrogate outcome studies)

Systematic review of well-designed cohort studies

Strong evidence from one cohort study

Unsystematic observations (expert opinions)

Topics Outline

• The Cycle of Research

• Understanding the Design of a Research Study
  – Sampling Methods
  – Study Designs
  – Measurement Scales
Measurement Scales

Categorical
- Ordinal
  - scale
- Nominal
  - not ordered

Numerical
- Continuous
  - any value
  - on the scale
- Discrete
  - integers
Categorical Scales

• Nominal
  – Examples: gender, race, blood type

• Ordinal
  – Ex. Apgar scores, tumor stage, social class
# Frequency Tables

<table>
<thead>
<tr>
<th></th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoker</td>
<td>100</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>25</td>
</tr>
<tr>
<td>Non-smoker</td>
<td>75</td>
</tr>
</tbody>
</table>
Contingency Tables

<table>
<thead>
<tr>
<th></th>
<th>Lung Cancer</th>
<th>No Lung Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoker</td>
<td>85</td>
<td>15</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Non-smoker</td>
<td>0</td>
<td>75</td>
</tr>
</tbody>
</table>
Numerical Scales

• Continuous
  – Examples: Blood pressure, temperature, age, weight, height, etc.

• Discrete
  – Examples: Number of children in a family, number of births in a year, number of accidents in a month, etc.
Relationships

• Numerical data can be converted to categorical data.

• Categorical data can be coded numerically.
Consider the Measurement Scale

• Categorical?
  – Nominal or Ordinal?
  – Binary?

• Numerical?
  – Discrete or Continuous?
Summary

• The Cycle of Research

• Understanding the Design of a Research Study
  – Sampling Methods
  – Study Designs
  – Measurement Scales