**Learning Objectives**

1. Describe the various designs of epi studies
2. Describe the key conceptual principles of design applications in infectious disease epidemiology

---

**Study design**

- Two fundamental tenets
  - Human diseases do not occur at random
  - Existence of causal and preventable factors
- Stochastic vs. Deterministic approaches
  - Stochastic: process with randomness
  - Deterministic: process by which event is determined by proportionality
- How you sample determines the study design and analysis

---

**Sampling Methods**

- Three distinct approaches
  - Probability sampling
    - Simple Random Sampling
    - Stratified Random Sampling
    - Multistage Sampling
  - Quasi probabilistic sampling
    - Systematic sampling—Nth name selection
  - Non-probability sampling
    - Convenience Sampling
    - Sequential sampling
    - Snowball sampling
**Sampling Scheme**

- Four schemes
  - Multinomial
    - Overall marginal total fixed *a priori*
  - Product multinomial
    - *One of the marginal totals fixed a priori*
  - Poisson
    - Fixed time determines all marginal totals
  - Finite population sampling (Hypergeometric)
    - Based on Ronald Fisher’s experiment
    - Only disproves a hypothesis but not endorse
    - Lady testing tea (1935)—Ms. Muriel Bristol

---

**Fundamentals of study design**

- What is fixed *a priori* determines the study

<table>
<thead>
<tr>
<th>Disease status</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure status</td>
<td>Yes</td>
<td>m₁</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>m₂</td>
</tr>
<tr>
<td>n₁</td>
<td>n₂</td>
<td>N</td>
</tr>
</tbody>
</table>

- Fix n₁ and n₂ first = Case-control study
- Fix m₁ and m₂ first = Cohort study
- Fix N first = Cross-sectional study
- Fix N(t) = Poisson process

---

**Requisites of ID Epi Study Designs**

- Three factors
  - Host
  - Agent
  - Environment

- Biologic variation in either the host or agent can influence the natural history of disease
- Environment can effect transmission
Basic distinctions of epi designs

- Three factors
  - Sampling method
    - By disease status—Case-control
    - By exposure status—Cohort
    - By fixing the total sample size—Cross-sectional
    - By event randomness/fixed time—Poisson
  - Timing of occurrence of disease/exposure
    - Time elapsing from exposure to outcome
  - Degree of “control” the design accords
    - Random allocation of subjects

Design of Epidemiological Studies

Common Strategies Employed

- Descriptive Studies
  - Factors describing people, place, time, activities
    - (Generate a Hypothesis)
- Analytical Studies
  - Establish association between Exposure (E) and Disease (D)
    - (Test a Hypothesis)

Choosing a study design

- Driven by the hypothesis
  - Smokers are more likely to develop lung ca.
  - Lung ca. patients are more likely to be smokers.
- Information available at investigation
  - Outcome is already there, exposure is to be investigated.
  - Exposure is putative or strongly suspected and outcome is to be observed.
  - Period of latency
Choosing study design (...cont.)
- Consider if disease or exposure is rare
  - Rare exposure with short latency – Cohort
  - Rare disease – Case-control
- Consider time needed for outcome to occur
  - Shorter term – Cohort, Experimental
  - Longer term – Case-control
- Amount of knowledge about a disease
  - Less known – Case-control
- Consider practicality
  - Sampling, Data Collection, Cost, and Analysis

Sampling scheme & design
- What is fixed a priori determines the study

The Basic Framework of Epidemiological Study Designs

Population at Risk

Prospective

Retrospective

Odds of E|D
\[ \frac{a}{d} \]

(Odds Ratio)

Case-Control

Cohort

Randomized Trial

Risk of Disease
\[ \frac{a(c+d)}{c(a+b)} \]

(OR: Odds Ratio)

Relative Risk
\[ \frac{a+c}{b+d} \]

Relative Risk
\[ \frac{a+c}{b+d} \]
Various designs
- Descriptive: Case Reports
- Ecological
- Cross-sectional
- Case-control
- Cohort
- Clinical trial (Experimental)
- Variations: nested studies, serial cross-sectional, etc.

Case Reports
- Rabies
- What was the significance of the two case reports presented?
  - 1) child and duck vaccine
  - 2) spelunker
- What should be the next step?

Case series
- AIDS cluster
  - Identification of men in NYC with Pneumocystis carinii pneumonia and men with Kaposi’s sarcoma in San Francisco
    - Pharmacy aid
  - What was the importance of this case series?
Ecological studies

Cross sectional studies

- Identification of disease and exposure at the same time
  - HAV and household characteristics on US/Mexico border
- Serial cross-sectional studies to determine trends
  - e.g., Newborn screening for HIV
**Case-control studies**
- Participants identified on the basis of disease status
- Outbreak investigations
  - Able to test for multiple associations
- Can also test a specific hypothesis
  - Good for rare diseases or diseases with a long latency period

**Cohort studies**
- Participants identified on the basis of exposure status
- Follow individuals over time
  - Prospective
  - Retrospective
- Best suited for rare exposures, shorter latency periods

**Types of Cohort Studies**

<table>
<thead>
<tr>
<th>Cohort Design</th>
<th>Time</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Past</td>
<td>Present</td>
<td>Future</td>
</tr>
<tr>
<td>1. Concurrent</td>
<td>E</td>
<td>E</td>
<td>D</td>
</tr>
<tr>
<td>(Prospective)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Nonconcurrent</td>
<td>E</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>a) Historical</td>
<td>E</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>(Retrospective)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Mixed</td>
<td>E</td>
<td>E</td>
<td>D</td>
</tr>
<tr>
<td>(Ambispective)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reference: Szklo, M & Nieto, F.J. Epidemiology beyond the basics, 2000
Gordis, L. Epidemiology, 3rd Ed. 2004
Schneider D. SuperCourse Lecture on Cohort Studies, UP
Clinical trials (Experimental)

- Generally used to evaluate a treatment or preventive measure, e.g. vaccination
- Randomization
- Blinding
- Most conclusive

Measures of association

<table>
<thead>
<tr>
<th>Design/Sampling</th>
<th>Measures of Association</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Relative Risk</td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔</td>
</tr>
<tr>
<td>Case-control</td>
<td>✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔</td>
</tr>
<tr>
<td>Prospective (Cohort)</td>
<td>✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔</td>
</tr>
</tbody>
</table>

Special study designs

- Transfusion-transmission of HIV
  - Novel way to determine viral differences
- Twin studies
  - Comparison of identical (monozygotic) vs fraternal (dizygotic)
- Genetic studies
- Vaccine probe
**Data collection**

- Identify and measure infection or disease
  - Biological marker, such as antibodies
    - Immunity waned vs. never existed
    - Exposed vs. vaccinated
    - Disease vs. infection
  - Case-definition based upon symptoms

**Data sources and limitations**

- Laboratory testing
- Symptom based case definition
- Personal interview
- Pre-existing data
  - Death certificates
  - ICD 9 Codes
  - Pathology reports

**Disease Identification and Data Source**

<table>
<thead>
<tr>
<th>Healthy</th>
<th>Disease Onset</th>
<th>Symptoms</th>
<th>Seek Care</th>
<th>Diagnosis</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Outcomes:
- Cure
- Control
- Disability
- Death

Sources of Data:
- Interviews
- Physician Records
- Hospital Records
Bias
- Selection bias
  - Hospital cases are more severe
  - Differences in enrollment between groups
- Information bias
  - Response-bias
  - Recall bias

Confounding
- A factor that distorts the apparent magnitude of the effect of a study factor on risk
  - AIDS and “poppers” in the 1980s—recreational drug use (amyl nitrite) and Kaposi Sarcoma
- Two forms of confounding
  - Database confounding
  - Literature-based confounding (Structural)

Causality (p348-9)
- Koch’s postulates—derived from an infectious disease model 1880s
  - Causative agent consistently isolated in ill
  - Not present in those without disease
  - Disease can be induced if organism is inoculated in a healthy individual
  - Organism can be grown in a lab
Bradford Hill Criteria for Causality

- Temporal association ↑↑↑
- Removing exposure results in reduction of disease (Experiment) ↑↑↑
- Dose-response relationship ↑
- Strength of association ↑
- Biologic plausibility ↑
- Consistency ↑
- (Specificity ↓, Analogy ↓, Coherence ↓)

Case-Cohort Studies
(Unmatched Nested Case-Control Study)
- Epi design that reconciles the benefits of cohort and case-control designs
- Time and cost efficient
- Control selected randomly from cohort at baseline ($t_0$) at $t_1$

Nested Case-Control Studies
(Incidence Density Sampling, Riskset)
- This also reconciles the benefits of cohort and case-control designs
- Time and cost efficient
- Control selected randomly from cohort at every event among the controls
Hierarchy of epi designs

<table>
<thead>
<tr>
<th>Design</th>
<th>Hypothesis Generation</th>
<th>Hypothesis Testing</th>
<th>Causal Inference</th>
<th>Potential Contextual Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case History</td>
<td>✓</td>
<td></td>
<td>Speculative</td>
<td>High</td>
</tr>
<tr>
<td>Aggregate</td>
<td>✓</td>
<td></td>
<td>Possibly Suggestive</td>
<td></td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>✓</td>
<td></td>
<td>Somewhat Suggestive</td>
<td></td>
</tr>
<tr>
<td>Case-control</td>
<td>✓</td>
<td>✓</td>
<td>Moderately Suggestive</td>
<td></td>
</tr>
<tr>
<td>Cohort Follow-up</td>
<td>✓</td>
<td>✓</td>
<td>Strongly Suggestive</td>
<td></td>
</tr>
<tr>
<td>Randomized Clinical Trial</td>
<td>✓</td>
<td>✓</td>
<td>Formative</td>
<td></td>
</tr>
</tbody>
</table>

Source: Adapted from Ibrahim, 1985

A Pearl of Wisdom

The present era of epidemiology is coming to a close. The focus on risk factors at the individual level—the hallmark of this era—will no longer serve. We need to be concerned equally with causal pathways at the societal level and with pathogenesis and causality at the molecular level.¹

Merryn Susser, MB, Bch, FRCP(E), DPH, and Ezra Susser, MD, DrPH

Am J Public Health 1994; 84:668-673