Descriptive and Analytic Studies

Created: 2013
Descriptive and Analytic Studies. Atlanta, GA: Centers for Disease Control and Prevention (CDC), 2013.
### Table of Contents

DESCRIPTIVE AND ANALYTIC STUDIES .............................................. 3

- **Learning Objectives** ........................................................................................................... 3
- **Estimated Completion Time** ................................................................................................. 3
- **References and Resources** ................................................................................................... 3
- **Module Content** .................................................................................................................... 5
- **Skill Assessment** .................................................................................................................. 30
Descriptive and Analytic Studies

**Learning Objectives**
At the end of the training, participants will be able to:

- Identify the following for an NCD problem:
  - Type of study to conduct
  - Sampling methods to use
  - Measure of association to calculate for a particular study
- Interpret the results of descriptive and analytic studies.

**Estimated Completion Time**

- 5 hours, 30 minutes (4 hours for interactive lecture; 1 hour, 30 minutes for Skill Assessment).

**References and Resources**

DESCRIPTIVE AND ANALYTIC STUDIES


Learning Objectives

- Identify the following for an NCD problem:
  - Type of study to conduct
  - Sampling methods to use
  - Measure of association to calculate for a particular study
  - Interpret the results of descriptive and analytic studies.

Lesson Overview

- Reasons for conducting studies
- Definition, characteristics, and analysis of:
  - Descriptive studies
  - Analytic studies
  - Methods of sampling

Why Conduct Studies?

To describe burden of disease or prevalence of risk factors, health behaviors, or other characteristics of a population that influences risk of disease
- To determine causes or risk factors for illness
- To determine relative effectiveness of interventions
Taxonomy of Epidemiologic Studies: Figure 1

Taxonomy of Epidemiologic Studies

Descriptive or Analytic Studies?

Descriptive studies
- Generate hypotheses
- Answer what, who, where, and when

Analytic studies
- Test hypotheses
- Answer why and how

Definition and Characteristics of Descriptive Studies
Descriptive and Analytic Studies

Descriptive Studies
Characterize who, where, or when in relation to what (outcome)
- **Person**: characteristics (age, sex, occupation) of the individuals affected by the outcome
- **Place**: geography (residence, work, hospital) of the affected individuals
- **Time**: when events (diagnosis, reporting, testing) occurred

Types of Descriptive Studies

- **Aggregate**
  - Ecological Studies
- **Individual**
  - Case Report
  - Case Series
  - Cross-sectional Study

Cross-Sectional Study as a Descriptive Study
**Purpose**: To learn about the characteristics of a population at one point in time (like a photo "snap shot")

**Design**: No comparison group

**Population**: All members of a small, defined group or a sample from a large group

**Results**: Produces estimates of the prevalence of the population characteristic of interest
### When to Conduct a Cross-Sectional Study

- To estimate prevalence of a health condition or prevalence of a behavior, risk factor, or potential for disease
- To learn about characteristics such as knowledge, attitude and practices of individuals in a population
- To monitor trends over time with serial cross-sectional studies

### Cross-Sectional Study Measures

**Prevalence of a condition:**

\[
\text{Prevalence} = \frac{\text{number of existing cases}}{\text{size of population}}
\]

(or population count)

### Example: Cross-Sectional Study

**Objective**

- To estimate the magnitude and patterns of violence against pregnant women

**Study**

- Population-based, household, cross-sectional study in Mboya and Dar es Salaam, Tanzania, 2001-2002

**Result**

- Violence experienced by 7% in Dar es Salaam and 12% in Mboya

*Note: Sample size and study design vary widely, so always consult a statistician or other expert in research methodology.*

*Reference: Tanzania Health Indicators 2012 (DHS 153E) 171-86.*
Studies to Track Trends in Newly Recognized Cases

Incidence study
- Newly reported or registered disease cases compared over time, place, or person
- Population estimates or other population group totals used as denominators

Ecological study
- Rates are linked to the level of exposure to some agent for the group as a whole

Example: Incidence Study

Objective
- To estimate the incidence and prevalence of diabetes in young persons in the United States

Study

Result
- Trends for diabetes death rates varied by age group

Example Incidence Study (continued)
**Analytic Studies Definition**

Analytic studies test hypotheses about exposure-outcome relationships

- Measure the association between exposure and outcome
- Include a comparison group

**Developing Hypotheses**

- A hypothesis is an educated guess about an association that is testable in a scientific investigation.
- Descriptive data (Who? What? Where? When?) provide information to develop hypotheses.
- Hypotheses tend to be broad initially and are then refined to have a narrower focus.
Developing Hypotheses Example
Hypothesis: People who smoke shisha are more likely to get lung cancer than people who do not smoke shisha.
- Exposure: smoking shisha
- Outcome: lung cancer

Hypothesis: ?
- Exposure: ?
- Outcome: ?

Analytic Study Types

<table>
<thead>
<tr>
<th>Experimental Studies</th>
<th>Observational Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized Control (Intervention) Trials</td>
<td>Cohort</td>
</tr>
<tr>
<td></td>
<td>Case-control</td>
</tr>
<tr>
<td></td>
<td>Cross-sectional</td>
</tr>
</tbody>
</table>

Cohort Studies

What is a cohort?
A well-defined group of individuals who share a common characteristic or experience
- Example: Individuals born in the same year

What are other examples of cohorts?
Cohort Study
(longitudinal study, follow-up study)
- Participants classified according to exposure status and followed-up over time to ascertain outcome
- Can be used to find multiple outcomes from a single exposure
- Appropriate for rare exposures or defined cohorts
- Ensures temporality (exposure occurs before observed outcome)

Cohort Study Design

Study Population

<table>
<thead>
<tr>
<th>Exposed</th>
<th>Unexposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>Disease</td>
</tr>
<tr>
<td>No Disease</td>
<td>No Disease</td>
</tr>
</tbody>
</table>

Exposure is self-selected, follow over time.

Types of Cohort Studies

Prospective cohort studies
- Group participants according to past or current exposure and follow-up into the future to determine if outcome occurs

Retrospective cohort studies
- At the time that the study is conducted, potential exposure and outcomes have already occurred in the past
### Prospective Cohort Studies
- **Study Population**
  - Exposed
    - Disease
    - No Disease
  - Unexposed
    - Disease
    - No Disease
- Start of study (Present) → (Future)

### Retrospective Cohort Studies
- **Study Population**
  - Exposed
    - Disease
    - No Disease
  - Unexposed
    - Disease
    - No Disease
- (Past) ← Start of study (Present)

### When to Conduct a Cohort Study
- When the exposure is rare and the outcome is common
  - Agricultural pesticide use and cancer events
- To learn about multiple outcomes due to a single exposure
  - Health effects of a nuclear power plant accident
Analysis of Cohort Studies

Risk:
Quantifies probability of experiencing the outcome of interest in a given population
- Calculation: Number of new occurrences of outcome/population at risk
Example:
- 29 new cases of diabetes in a community
- 100,000 people in the community at risk for diabetes
- What is the risk of diabetes? 29/100,000

Analysis of Cohort Studies: Person-Time, Rate

Quantifies occurrence of outcome in population by time
Calculation: number of new cases during follow-up period / Sum of time each study participant was followed and at risk of disease
Example: 1,212 tunnel workers
160 deaths among tunnel workers 24,035 person-years at risk
Mortality rate = 160 / 24,035 = 6.7 deaths per 1,000 workers per year

Risk Ratio

- Can also be called Relative Risk or RR
- Quantifies a population’s risk of disease from a particular exposure
- Calculation:
  Risk in the exposed group / Risk in the unexposed group
Rate Ratio

Compares the rates of disease in two groups that differ by demographic characteristics or exposure history.

Calculation:

\[
\frac{\text{Rate for group of primary interest}}{\text{Rate for comparison group}}
\]

RR Strength Scales

<table>
<thead>
<tr>
<th>RR</th>
<th>Strength</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.71 – 0.99</td>
<td>Weak</td>
<td>1.01 – 1.50</td>
</tr>
<tr>
<td>0.41 – 0.70</td>
<td>Moderate</td>
<td>1.51 – 3.00</td>
</tr>
<tr>
<td>0.00 – 0.40</td>
<td>Very strong</td>
<td>&gt;3.00</td>
</tr>
</tbody>
</table>

Example: Risk Ratio

**Question:** What is the relationship between being obese and getting type 2 diabetes?

\[
\text{Risk Ratio} = \frac{0.0076}{0.0013} = 5.8
\]

**Interpretation:** The risk of diabetes among those who are obese is 5.8 times the risk among those who are not obese.
Example: Person-Time Rate Ratio

**HANES — Follow-up Study (male diabetics subset)**
- Original enrolment: 1971–1975
- Follow-up: 1982–1984
- Complete follow-up on:

<table>
<thead>
<tr>
<th></th>
<th>Enrolled</th>
<th>Died</th>
<th>PY of FU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetics</td>
<td>189</td>
<td>160</td>
<td>1414.7</td>
</tr>
<tr>
<td>Non-Diabetics</td>
<td>3151</td>
<td>811</td>
<td>28,028.8</td>
</tr>
</tbody>
</table>

- Mortality Rate Ratio:
  - \( \frac{160}{1414.7} \times \frac{28029.8}{7677000} = 2.5 \)

---

Case-Control Study

**Purpose:**
- To study rare diseases
- To study multiple exposures that may be related to a single outcome

**Study Subjects**
- Participants selected based on outcome status:
  - Case-subjects have outcome of interest
  - Control-subjects do not have outcome of interest

---

Case-Control Study Design

- **Exposed**
  - Assess exposure history
  - Identify cases and select controls
- **Unexposed**
  - Cases (Diseased)
- **Exposed**
  - Controls (No Disease)
- Source Population
**When to Conduct a Case-Control Study**

- The outcome of interest is rare
- Multiple exposures may be associated with a single outcome
- Funding or time is limited

**Case-Control Study: Analysis Format**

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Exposure odds ratio (OR) = RR when disease is rare
Odds of being exposed among the cases = a/c
Odds of being exposed among the controls = b/d
Exposure odds ratio = \((a/c)/(b/d) = (a*d)/(b*c)\)
(Cross-product ratio)

**Example
Odds Ratio**

<table>
<thead>
<tr>
<th>Work in mine?</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>No</td>
<td>83</td>
<td>97</td>
</tr>
</tbody>
</table>

Odds Ratio = \(17/83 + 13/87 = 1.37\)
**Prevalence Ratio and Prevalence Odds Ratio**

- Chronic disease – date of onset is unknown
- Measure prevalence rather than incidence

<table>
<thead>
<tr>
<th>RR</th>
<th>PR (prevalence ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>POR (prevalence odds ratio)</td>
</tr>
</tbody>
</table>

**Prevalence Ratio**

- Usually from a cross-sectional study
- Similar to risk ratio from cohort study

<table>
<thead>
<tr>
<th>Exposure</th>
<th>With disease</th>
<th>Without disease</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>Unexposed</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></td>
</tr>
</tbody>
</table>

- PR = Prevalence of disease in exposed group / Prevalence of disease in unexposed group
- OR = a/(a+b) / c/(c+d)

**Prevalence Odds Ratio**

- Usually from a cross-sectional study
- Similar to odds ratio from case control study
- Calculated same way as odds ratio:

\[ \text{POR} = \frac{a*d}{c*b} \]

<table>
<thead>
<tr>
<th>With disease</th>
<th>Without disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
</tr>
<tr>
<td>Unexposed</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
</tr>
</tbody>
</table>
Example: Prevalence Ratio and Prevalence Odds Ratio

Prevalence of Breast Cysts

<table>
<thead>
<tr>
<th>Lifetime use of oral contraceptives</th>
<th>Yes Cyst</th>
<th>No Cyst</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ever Used</td>
<td>124</td>
<td>3123</td>
<td>3247</td>
</tr>
<tr>
<td>Never Used</td>
<td>77</td>
<td>2557</td>
<td>2634</td>
</tr>
<tr>
<td>Total</td>
<td>201</td>
<td>5680</td>
<td>5881</td>
</tr>
</tbody>
</table>

Prevalence of breast cysts among ever users = 124/3247 = 0.038
Prevalence of breast cysts among never users = 77/2634 = 0.029
Prevalence ratio = 0.038/0.029 = 1.3
Prevalence odds ratio = 124*2634/77*3247 = 1.3

Practice Exercise #1

Background:
- NCDs such as type 2 diabetes are poorly understood and under-prioritized in many low-to-middle income countries.
- You want to determine the risk of type 2 diabetes associated with cardiovascular risk factors such as obesity and abdominal fat mass in your country.

Questions:
1. What type of study would you conduct and why?
2. What is the measure of association to calculate for this study?

Practice Exercise #2

Background:
- The prevalence of prostate cancer has increased in your country over the last 5 years.
- You want to examine the association between calcium intake and prostate cancer risk.
- You have limited time and funding to conduct this study.

Questions:
1. What type of study would you conduct and why?
2. What is the measure of association to calculate for this study?
Practice Exercise #3

**Background:**
- Cardiovascular disease (CVD) is of growing concern; however your country has no recent data on the burden of this disease.
- You want to estimate the burden of cardiovascular disease in the two main cities in your country.

**Questions:**
1. What type of study would you conduct and why?
2. What is the measure of association to calculate for this study?

---

**METHODS OF SAMPLING**

---

**Discussion Question**

*Why do we use sampling?*

- Cannot get information on everyone in a population
- Efficiently gets information on a large population
- Obtains a representative sample of a population
Sampling Methods
Two main types of sampling methods:
• Probability sampling
• Non-probability sampling

Probability Sampling
What are types of probability-based samples?
• Simple random sampling
• Systematic random sampling
• Stratified random sampling
• Cluster sampling

Simple Random Sample
Principle
• Equal chance/probability of drawing each unit

Procedure
• List all units (persons) in a population
• Assign a number to each unit
• Randomly select units
Method: Simple Random Sampling

Each unit has the same probability of selection (1/30)

---

Example: Simple Random Sample
Example: Calculate the prevalence of tooth decay among 1200 children attending a school

(sample size = 100)

- List all children attending the school
- Each child assigned a number from 1 to 1200
- Randomly select 100 numbers between 1 and 1200

---

Advantages & Disadvantages: Simple Random Sample

Advantages
- Simple

Disadvantages
- Need complete list of units
- Units may be scattered and poorly accessible
**Systematic Random Sample**

**Principle**
- Select sample at regular intervals based on sampling fraction

**Procedure**
- List all units (persons) in a population
- Assign a number to each unit
- Calculate sampling fraction (population size / sample size)
- Select first unit at random based on sampling fraction
- Subsequent units are chosen at equal intervals

---

**Advantages & Disadvantages: Systematic Random Sample**

**Advantages**
- Simple
- Can be implemented easily without software

**Disadvantages**
- Need complete list of units

---

**Example: Systematic Random Sample**

*Example: Calculate the prevalence of tooth decay among 1200 children attending a school (sample size = 100)*
- List all children attending the school
- Randomize the list to avoid bias
- Each child assigned a number from 1 to 1200
- Sampling fraction = 1200/100 = 12
- Randomly select a number between 1 and 12
  - Example: 8
  - Select every 12th child, starting with child #8
    - Example: 8, 20, 32, 44…
**Stratified Random Sample**

**Principle**
- Select random samples from within homogeneous subgroups (strata)

**Procedure**
- List all units (persons) in a population
- Divide the units into groups (called strata)
- Assign a number to each unit within each stratum
- Select a random sample from each stratum
- Combine the strata samples to form the full sample

---

**Method: Stratified Random Sample**

- Sampling frame divided into groups (age, sex, socioeconomic status)
- Units in each group have the same probability of selection, but probability differs between groups

<table>
<thead>
<tr>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Men Sample" /></td>
<td><img src="image" alt="Women Sample" /></td>
</tr>
</tbody>
</table>

Probability: 1/20  Probability: 1/15

---

**Advantages & Disadvantages: Stratified Random Sample**

**Advantages**
- Can get separate estimates from the whole population and from individual strata (if sample is large enough)
- Precision increased if less variability within strata than between strata

**Disadvantages**
- Can be difficult to identify strata
Class Discussion Question
What are some examples of strata that you might sample within?
- Race/ethnicity/tribe/nationality
- Age group
- Gender
- Geographic location
- Socioeconomic status
- Smoking status
- Occupation
- Education
- Many possibilities!

Example: Stratified Random Sample
Example: Calculate the prevalence of tooth decay among 1200 children attending a school, with equal representation of males and females (sample size = 100)
- List all children attending the school
- Divide the children into two groups
  - 546 males and 660 females
- Assign each child a number
  - Males 1 to 540
  - Females 1 to 660
- Randomly select 50 males and 50 females

Cluster Sample
Principle
- Select all units within randomly selected geographic clusters
Procedure
- Divide population into geographic groups (clusters)
- Assign a number to each cluster
- Randomly select clusters
- Sample all units within selected clusters OR select a random sample of units within selected clusters
Advantages & Disadvantages: Cluster Sample

Advantages
- List of sampling units not required
- More efficient for face-to-face interviews when units are dispersed over a large area

Disadvantages
- Loss of precision due to correlation within clusters
- This correlation needs to be taken into account in sample size calculations and analysis ("design effect")

Non-probability Sampling

- Probability of selection is unknown or zero
- Inexpensive
- Results not generalizable
- Results often biased

Common types of non-probability sampling:
- Convenience sampling
- Snowball sampling / Respondent-driven sampling
- Voluntary sampling

Choosing a Sampling Method

Consider:
- Population to be studied
  - Size/Geographic distribution
  - Availability of list of units
  - Heterogeneity with respect to variable
- Level of precision required
- Resources available
Practice Exercise #4

**Background:** You will choose a sampling method for each of the following studies.

**Questions:**

What sampling method would you use for:

1. The cross-sectional study on CVD described in Practice Exercise #3? Why?
2. A one-time survey of citizens' attitudes toward smoking and second-hand smoke in response to proposed legislation to impose a ban on smoking in restaurants. Why?
3. Serosurvey of blood lead levels (or urinary arsenic levels) of prisoners entering the nation's largest prison (or pregnant women entering the nation's largest maternity ward) to determine average level of exposure in the population.

---

**SUMMARY**

---

**Descriptive vs. Analytic Epidemiology**

Descriptive epidemiology:
- Who, What, When, and Where

Analytic epidemiology:
- Why and How
Types of Descriptive and Analytic Studies

Types of descriptive studies
- Aggregate: Ecological study
- Individual: Case report, case series, cross-sectional study

Types of analytic studies
- Experimental: Randomized control trial
- Observational: Cohort, case-control, cross-sectional

Cohort vs. Case-Control Studies

<table>
<thead>
<tr>
<th>Study Comparison</th>
<th>Cohort Study</th>
<th>Case-Control Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferred Study Design When...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Members are easily identified</td>
<td></td>
<td>Identifying entire cohort would be too costly or time consuming</td>
</tr>
<tr>
<td>Members are easily accessible</td>
<td></td>
<td>Accessing entire cohort would be too costly or time consuming</td>
</tr>
<tr>
<td>Exposure is rare</td>
<td></td>
<td>Illness is rare</td>
</tr>
<tr>
<td>There may be multiple diseases involved</td>
<td></td>
<td>There may be multiple exposures involved</td>
</tr>
<tr>
<td>Study Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposed persons</td>
<td></td>
<td>Persons with illness (cases)</td>
</tr>
<tr>
<td>Comparison Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexposed persons</td>
<td></td>
<td>Persons without illness (controls)</td>
</tr>
</tbody>
</table>

Sampling Advantages and Disadvantages

<table>
<thead>
<tr>
<th>Probability Sampling</th>
<th>Non-Probability Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advantages</td>
<td>Advantages</td>
</tr>
<tr>
<td>Results are generalizable</td>
<td>Easy</td>
</tr>
<tr>
<td>Representative</td>
<td>Quick access to certain groups</td>
</tr>
<tr>
<td>Disadvantages</td>
<td>Disadvantages</td>
</tr>
<tr>
<td>Expensive</td>
<td>Not representative</td>
</tr>
<tr>
<td>Logistically difficult</td>
<td>Results are not generalizable</td>
</tr>
<tr>
<td>Time-intensive</td>
<td></td>
</tr>
</tbody>
</table>
Skill Assessment

- You will work in small groups to complete two parts of a skill assessment:
  1. Identify the type of study to conduct and sampling method
  2. Interpret the results
- Materials and questions can be found in your Participant Guide.
- Spend approximately 1 hour completing the assessment.
- Be prepared to share the group’s work.
Instructions:

1. You will work in small groups to complete a three-part assessment.
2. Select a member of your group to record your responses.
3. Groups will have approximately 1 hour to complete the assessment.
4. At the end of the assessment, one member from the group will share your summary with the class. (20 minutes)

Part 1. Identify study to conduct and sampling method to use (30 minutes)

Read the following three issues of public health concern and answer the questions that follow.

1. Cardiovascular disease is an increasing problem in Tanzania. The last study to examine risk factors of cardiovascular disease in Tanzania was in 1987. You want to know about the current status of risk factors for cardiovascular disease and would like to examine how the risk factors have changed since the last study in 1987.
   a. What type of study would you conduct and why?
   b. What sampling method would you use?

2. In Thailand, breast cancer incidence is increasing, but little is known about the primary risk factors for breast cancer among Thai women. Current understanding of breast cancer risk factors are from studies in high-income countries. It is unclear if identified risk factors in high-income countries are the same in Thailand. You want to learn about the risk factors for breast cancer in Thailand females.
   a. What type of study would you conduct and why?
   b. How would you use simple random sampling to select 860 controls?
   c. How would you use systematic random sampling to select 860 controls?
   d. How would you use stratified random sampling to select 860 controls?

3. Non-smoking women in China are disproportionately exposed to environmental tobacco smoke exposure in their homes and at work. You want to know if non-smoking Chinese women exposed to environmental tobacco smoke have increased mortality compared to non-smoking Chinese women who live and work in smoke-free
environments. You’d also like to know if environmental tobacco smoke exposure has an impact on the risk of cardiovascular and cancer deaths among Chinese women.

a. What type of study would you conduct and why?

Part 2. Interpreting the results (30 minutes)
Interpret the results for each study in 1-2 concise sentences. If applicable, explain whether the exposure was associated with an increased or decreased chance of experiencing the outcome.

1. You found that the prevalence of hypertension was 41.1% in men and 38.7% in women.

2. You found that women with older siblings had a statistically significant odds ratio of 0.3 for breast cancer.

3. You found that non-smoking women whose husbands smoked had a statistically significant risk ratio of 1.19 for mortality due to cancer.