Introduction to Study Design

Dr. Mary E. Hovinga

Descriptive Characteristics

Person
Age, Sex, Race, Ethnicity, Marital Status, Parity, Occupation, Education, SES, Alcohol, Smoking, Diet, Genetic Markers

Place
Does disease frequency differ by:
• Country, State, County, City? (political units)
• Rural vs. Urban, Altitude, Latitude, Rainfall, Sunlight, Distance from the hazardous waste site? (physical units)

Time
1) Secular – Trends over years (can be difficult to interpret)
2) Cyclical – Seasonal trends (infectious diseases)
3) Short-term changes - Epidemic

Common Sequence of Studies
Case Reports
Recognizing a problem

Descriptive Studies
Generate clues about what is going on. Possible hypotheses for: Exposure → Disease
(Cross-Sectional, Ecologic)

Analytic Studies
Testing a specific hypothesis about E → D
(Case-Control, Cohort)

Cross-Sectional Study (survey)
A study which measures the disease and exposure status as they exist in a defined population at one specific point in time.
Cross-Sectional Study
A “snapshot” of:
• the current health status of the population (can identify high prevalence subgroups)
• the current exposure/behaviors of the population (can identify high exposure subgroups)


Prevalence = \( \frac{\text{Number of EXISTING cases}}{\text{Total Population}} \)

Incidence = \( \frac{\# \text{ NEW cases occurring in a time period}}{\text{Total Population at Risk}} \)

Asking the Question(s):
• Does this “exposed” community have more cancer, asthma, learning disabilities, hypertension than an “unexposed” community? (External comparison)

• Within this community, do people who are “exposed” have more cancer, asthma, etc than do people who are not exposed? (Internal comparison)

Advantages of Cross-Sectional Studies
• Tells you what is in the population right now (level and types of exposure/behaviors, illness), and can be done right now.

• Can look at health status, conditions, symptoms which are not routinely reported (in medical records)

• Useful for diseases which have no clear point of onset, or are recurring (measuring incidence is a problem)

Problems with Cross-Sectional Studies
• Measuring Prevalence – depends on Incidence and Duration

• Cannot demonstrate temporal link:
  - that exposure happened before disease
  - no time-trend analysis

Case Reports
Recognizing a problem

Descriptive Studies
(Cross-Sectional, Ecologic)
Generate clues about what is going on. Possible hypotheses for: Exposure —— Disease

Analytic Studies
(Case-Control, Cohort)
Testing a specific hypothesis about E —— D
Exposure → Disease

Hypothesis: The contaminated well water is associated with childhood leukemia

A statistical relationship, not necessarily a causal one.

Exposure → Disease

Can test by two different designs:

1. Observe a group of exposed children and a group of unexposed children for 2, 5, 10 years and compare how many in each group develop leukemia.

   **COHORT**

   (measuring and comparing Incidence of leukemia)

   • To measure the association between the exposure and the disease, compare $I_E$ and $I_U$

     \[
     RR = \frac{I_E}{I_U} = \frac{\text{Incidence in the exposed}}{\text{Incidence in the unexposed}}
     \]

     \[
     \begin{array}{cc}
     & D \quad \overline{D} \\
     E & a & b \\
     \overline{E} & c & d \\
     \end{array}
     \]

     \[
     RR = \frac{a}{a+b} \div \frac{c}{c+d}
     \]

     RR = Relative Risk (Relative Rate)  
     Risk Ratio (Rate Ratio)

   Interpreting RR

   \[
   RR = 3.0 \quad \text{(positive association)}
   \]

   • The exposed are 3 times as likely to develop the disease as are the unexposed.

   \[
   RR = 1.0 \quad \text{(no association)}
   \]

   • The exposed are 1.0 times as likely (equally likely) to develop the disease as are the unexposed.

   \[
   RR = 0.5 \quad \text{(negative association or protective)}
   \]

   • The exposed are 0.5 times as likely (half as likely) to develop the disease as are the unexposed.

Retrospective Cohort Study

• An Exposure in search of a Disease

• A study in which both the exposure and the disease have already occurred.

• Study subjects are still identified and enrolled based on their exposure status.

• Disease is determined only after the exposure status is classified.

• A common design in occupational epidemiology or in similar circumstances where exposure and medical records are available.

• Availability of records is often the limiting condition. Environmental situations – often use birth records, death certificates.

• Calculate SMR – the Standardized Mortality Ratio

• Compares OBSERVED events versus EXPECTED events
Exposure → Disease

2. Identify a group of children with leukemia and a group without leukemia and compare how many in each group drank contaminated well water.

**CASE-CONTROL**
(comparing the proportion of exposure)

---

**Case-Control Study**

- Selecting people with disease (cases) and people without the disease (controls) and comparing their exposure status.

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>a</td>
</tr>
<tr>
<td>E</td>
<td>c</td>
</tr>
<tr>
<td>b</td>
<td>d</td>
</tr>
</tbody>
</table>

**Odds Ratio**

- The measure of association in a case-control study is the Odds Ratio.

\[
OR = \frac{\text{odds of exposure among diseased}}{\text{odds of exposure among non-diseased}}
\]

\[
\begin{array}{c|c|c}
 & a & b \\
\hline
E & c & d \\
\end{array}
\]

\[
OR = \frac{a}{c} = \frac{ad}{bc}
\]

**Interpreting Odds Ratio**

OR = 3.0 (positive association)
- The cases are 3 times as likely to have the exposure as the controls.

OR = 1.0 (no association)
- The cases are 1.0 times as likely (equally likely) to have the exposure as the controls.

OR = 0.5 (negative association)
- The cases are 0.5 times as likely (half as likely) to have the exposure as the controls.

---

**Case-Control Studies**

**Advantages**
- Quick
- Relatively inexpensive
- Good for rare diseases
- Can study many exposures
- No drop-outs

**Disadvantages**
- Cannot measure incidence
- Specific to one disease
- Bad for rare exposures
- Temporal link unclear
- Vulnerable to bias

---

**Odds Ratio**

- The Odds Ratio is also the measure of association in a Cross-Sectional study. Becomes a Prevalence Odds Ratio.

\[
\begin{array}{c|c|c}
 & a & b \\
\hline
E & c & d \\
\end{array}
\]

\[
OR = \frac{a}{c} = \frac{ad}{bc}
\]
Judging OR’s and RR’s
(a guide, not a commandment)

Odds Ratio, Rel Risk = 1.0    No Association
OR, RR  1.1  -  2.0  Weak Association
       2.1  -  4.0  Moderate Assoc
          > 4.0  Strong Association