Cohort Studies

Sukon Kanchanaraksa, PhD
Johns Hopkins University
Design of a Cohort Study

Identify:
- Exposed
- Not exposed

Follow:
- Develop disease
- Do not develop disease
- Develop disease
- Do not develop disease
Cohort Study

First, identify

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not exposed</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a + b</td>
<td>c + d</td>
<td></td>
</tr>
</tbody>
</table>

Cohort Study

Then, follow to see whether

First, identify

<table>
<thead>
<tr>
<th></th>
<th>Disease develops</th>
<th>Disease does not develop</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>Not exposed</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
</tbody>
</table>
Cohort Study

Then, follow to see whether Disease develops and compare Disease does not develop Totals Incidence of disease

<table>
<thead>
<tr>
<th></th>
<th>Disease develops</th>
<th>Disease does not develop</th>
<th>Totals</th>
<th>Incidence of disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
<td>( \frac{a}{a + b} )</td>
</tr>
<tr>
<td>Not exposed</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
<td>( \frac{c}{c + d} )</td>
</tr>
</tbody>
</table>

\[
\frac{a}{a + b} = \text{Incidence in exposed} \quad \frac{c}{c + d} = \text{Incidence in not exposed}
\]
**Cohort Study**

Then, follow to see whether

Calculate

<table>
<thead>
<tr>
<th></th>
<th>Develop CHD</th>
<th>Do not develop CHD</th>
<th>Totals</th>
<th>Incidence of disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke cigarettes</td>
<td>84</td>
<td>2916</td>
<td>3000</td>
<td>84/3000</td>
</tr>
<tr>
<td>Do not smoke cigarettes</td>
<td>87</td>
<td>4913</td>
<td>5000</td>
<td>87/5000</td>
</tr>
</tbody>
</table>

\[
\frac{84}{3000} = 0.028 = \text{Incidence in 'smoke cigarettes'} \\
\frac{87}{5000} = 0.0174 = \text{Incidence in 'not smoke cigarettes'}
\]
Design of a Cohort Study

Begin with: Defined population

Non-randomized

Identify:
- Exposed
- Not exposed

follow:
- Develop disease
- Do not develop disease
- Develop disease
- Do not develop disease
Comparison of Experimental vs. Observational Study

**Experimental**
(Randomized trial)

- Population
  - Random allocation
    - Group A
    - Group B

**Observational**
(Cohort study)

- Population
  - Other-than-random allocation
    - Group A
    - Group B
Types of Cohort Studies

- **Prospective cohort study**
  - Concurrent cohort study or longitudinal study
- **Retrospective cohort study**
  - Non-concurrent cohort or historical cohort study

= Investigator
Time Frames for a Hypothetical Prospective Cohort Study Conducted in 2000

Prospective

2000

Defined population

Non-randomized

Exposed

Disease

No disease

Not exposed

Disease

No disease
Time Frames for a Hypothetical Retrospective Cohort Study Conducted in 2000

Defined population

Non-randomized

Exposed

Disease

No disease

Not exposed

Disease

No disease

Retrospective

1980

2000
Differentiating between Prospective and Retrospective

- **Prospective cohort study**
  - Investigator
    - Starts the study (from the beginning) with the identification of the population and the exposure status (exposed/not exposed groups)
    - Follows them (over time) for the development of disease
    - Takes a relatively long time to complete the study (as long as the length of the study)
Differentiating between Prospective and Retrospective

Retrospective cohort study

Investigator

- Uses existing data collected in the past to identify the population and the exposure status (exposed/not exposed groups)
- Determines at present the (development) status of disease

Investigator spends a relatively short time to:

- Assemble study population (and the exposed/not exposed groups) from past data
- Determine disease status at the present time (no future follow-up)
Investigator uses existing data collected in the past to:

- Identify the population and the exposure status (exposed/not exposed groups)
- **Follow them into the future** for the development of the disease

Investigator

- Spends a relatively short time to assemble study population (and the exposed/not exposed groups) from past data
- Will spend additional time following them into the future for the development of disease
Example of a Prospective Cohort Study: Framingham Study

- Defined Framingham study population
  - Exposed
    - Disease
    - No disease
  - Not exposed
    - Disease
    - No disease

Prospective

1948
Framingham Study

Objectives
To study the impact of several factors on incidence of cardiovascular diseases

Exposures
Blood pressure, smoking, body weight, diabetes, exercise, etc.

Multiple Outcomes
Coronary heart disease, stroke, congestive heart failure, peripheral arterial disease
Framingham Study as a Cohort Study

- The study started with a defined population
  - Investigators (USPHS and NHLBI) started by identifying a new population and did not use existing data to identify the population and the exposure groups
- There were several hypotheses to be tested
  - Different exposures and different outcomes
- For each exposure, investigators identified the “exposed” and the “not exposed” groups
- For each exposure, the participants were followed for the development of disease
- Different exposures were studied, as well as different diseases
**Derivation of the Framingham Study Population**

<table>
<thead>
<tr>
<th>Category</th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sample</td>
<td>3074</td>
<td>3433</td>
<td>6507</td>
</tr>
<tr>
<td>Respondents</td>
<td>2024</td>
<td>2445</td>
<td>4469</td>
</tr>
<tr>
<td>Volunteers</td>
<td>312</td>
<td>428</td>
<td>740</td>
</tr>
<tr>
<td>Respondents free of CHD*</td>
<td>1975</td>
<td>2418</td>
<td>4393</td>
</tr>
<tr>
<td>Volunteers free of CHD</td>
<td>307</td>
<td>427</td>
<td>734</td>
</tr>
<tr>
<td>Total Free of CHD</td>
<td>2282</td>
<td>2845</td>
<td>5127</td>
</tr>
</tbody>
</table>

*CHD = coronary heart disease*
Follow-Up of Participants

- Risk factors and the development of cardiovascular events were evaluated every two years by medical history, medical record review, and physical examination.
- All diagnoses were verified without knowledge of risk factors by Framingham examiners who reviewed medical records and death certificates.
- Approximately three percent of the subjects were lost to follow-up for mortality during the first 45 years of the study.
Timeline of Milestones from the Framingham Study

- **1948**: start of the Framingham Heart Study
- **1960**: cigarette smoking found to increase risk of heart disease
- **1961**: cholesterol, blood pressure, and ECG abnormalities found to increase risk of heart disease
- **1965**: first Framingham Heart Study report on stroke
- **1967**: physical activity found to reduce risk of heart disease; obesity to increase the risk
- **1970**: high blood pressure found to increase the risk of stroke
- **1974**: diabetes found to be associated with cardiovascular disease
- More milestones: [www.framingham.com/heart/timeline.htm](http://www.framingham.com/heart/timeline.htm)
A Change of Heart: How the People of Framingham, Massachusetts, Helped Unravel the Mysteries of Cardiovascular Disease
Average Annual Incidence of Coronary Heart Disease by Weight, Gender, and Age Group

- **Above median weight**
- **Below median weight**

### Men
- Age group: 40-49, 50-59, 60-69, 70-79
- Rate per 1,000

### Women
- Age group: 40-49, 50-59, 60-69, 70-79
- Rate per 1,000
Average Annual Incidence of Coronary Heart Disease by Systolic Blood Pressure
CHD Risk Assessment Based on Relationship Between HDL and LDL Cholesterol Men 50–70 Years

Types of Potential Bias in Cohort Studies

- **Selection bias**
  - Select participants into exposed and not exposed groups based on some characteristics that may affect the outcome

- **Information bias**
  - Collect different quality and extent of information from exposed and not exposed groups
  - Loss to follow-up differs between exposed and not exposed (or between disease and no disease)

- **Misclassification bias**
  - Misclassify exposure status or disease status
When Is a Cohort Study Warranted?

- When the (alleged) exposure is known
- When exposure is rare and incidence of disease among exposed is high (even if the exposure is rare, determined investigators will identify exposed individuals)
- When the time between exposure and disease is relatively short
- When adequate funding is available
- When the investigator has a long life expectancy
What are the differences in the study design between prospective cohort study and retrospective cohort study?

What are the differences in the study design between randomized clinical trial study and cohort study?

Why is cohort study preferred for studying rare exposure?