1. Why do an Observational Study?

An observational study investigates relationships between variables and searches for causal explanations.

But “treatments” are not assigned at random.

Investigators have no control over how the treatments are assigned.

For this reason, there is a strong possibility that relationships between explanatory and response variables may be distorted by confounding.

- Sometimes a randomized experiment is impossible because the explanatory variable is a characteristic that cannot be easily manipulated or assigned
  
  Example: sex, poverty status, tolerance for exercise
- Sometimes a randomized trial is unethical because one or more treatments is too harmful to give to human subjects
  
  Example: asbestos and lung cancer
- Sometimes a randomized trial is unethical because it would withhold effective treatment from a control group
  
  Example: giving a placebo to the seriously ill

Why do an observational study?

For making causal conclusions, observational studies are generally considered to be inferior to randomized experiments.

Nevertheless, there are some important reasons why we do observational studies.

Reason 1:

A randomized trial may be impossible, impractical or unethical.

Reason 2:

With observational studies, we may explore the effects of many explanatory variables at once.

With a randomized experiment, the treatment is specified in advance. We can look for many “effects,” but there can be only one cause.

With an observational study, we have the chance to identify many risk factors that may simultaneously contribute to the condition.

Example: Observational studies have shown that teenage alcohol use is strongly related to positive expectancies, but only weakly related to negative expectancies
Reason 3:

**Observational studies may have greater external validity.**

External validity means that the conclusions from a study can be generalized to a broader context.

- Subjects in randomized trials often chosen for convenience, accessibility or other reasons and may not be typical of the intended population
- Some observational studies use truly representative samples
- Observational studies may give a better indication of how treatments may affect subjects in their natural surroundings

2. Types of Observational Studies

Section 5.4 mentions two types of observational studies: case-control and prospective. But there is another kind that is extremely common.

A. Survey

A survey is any study that collects information from a sample of units and tries to generalize to a larger population.

But a survey can also serve as an observational study. For this purpose, it helps if the survey is longitudinal (i.e. collecting information on the same subjects at two or more occasions in time)

Example

Add Health is a nationally representative, longitudinal survey of American high-school students.

In Wave 1, students were asked a variety of questions on health and health-related behaviors.

In Wave 2, students were asked the same questions one year later.

Using these data, one can address questions like this: “Does dieting and/or exercise at Wave 1 have any relationship to change in body-mass index (BMI) between Waves 1 and 2?”

Warning:

When using a survey as an observational study, the subjects who are receiving one “treatment” (e.g. dieting) may be very different from those who receive another treatment (e.g. no dieting).

This means that there could be many possible confounders, and the confounding effects could be strong.

Other types of observational studies may do a better job at reducing confounding, because they may allow you to select groups that are more similar.

B. Case-control study

In a case-control study, we first identify subjects who have a particular outcome (e.g. pancreatic cancer). These are the cases.

Then we try to find a group of subjects who, in many respects look similar to them but do not have that outcome. These are called the controls.

Then we collect as much historical information as we can to see if these two groups are different in any other ways (e.g. amount of coffee consumption). The differences that we find may be regarded as potential causes. But causality is not proven, because it’s an observational study.

A case-control study is

- **Backward**: Subjects are selected into the study on the basis of the response variable (whether they have the condition or not), not the explanatory variable
- **Retrospective**: The data that are collected are based on the subjects’ memories of what they experienced in the (often distant) past

Warning: these retrospective reports are often flawed, subject to uncertainty and recall bias
Why do a case-control study?

- Sample size: If we are investigating a rare condition, we can get enough subjects who have the condition ("cases") without taking an enormous sample from the population.
- Speed: By finding subjects who already have the condition, we can do the study quickly; we don’t have to wait a long time until a reasonable number of subjects develop the condition.

C. Prospective observational study

A prospective observational study moves forward in time, not backward.

In a prospective observational study, we select subjects on the basis of the explanatory variable, not the response variable. Then we wait for the response to develop.

Example

To see if drinking coffee increases the risk of pancreatic cancer, we could find a group of people who drink coffee, and then find a group of people who look “similar” but do not drink coffee. Then we follow them for a long time to see if there is any difference in rate of pancreatic cancer between the two groups.

Advantages of a prospective study

- Better data: Following subjects in time and recording what happens tends to give more accurate data than retrospective reports.
- Greater reliability: Scientists generally regard a prospective study as more trustworthy than a retrospective one.

Disadvantages

- May take a long time.
- May need to start with a very large sample in order to eventually have enough subjects who eventually develop the condition.

Example

Researchers suspect that smoking during pregnancy may be related to low birth weight and infant mortality.

In designing a study, do you think it would be better to use a prospective or retrospective approach?

Answer

In this case, low birth weight is evident immediately when the baby is born. Infant mortality occurs during the first year. Therefore, one can do this study prospectively and see the results in less than two years.

Example

Researchers suspect that frequent use of anti-inflammatory drugs (e.g., Tylenol) may reduce the risk of Alzheimer’s disease among the elderly.

In designing a study, do you think it would be better to use a prospective or retrospective approach?

Answer

A prospective study might take a very long time (e.g., twenty or thirty years or more). A retrospective case-control study would be much cheaper and faster.