"You can't fix by analysis what you bungled by design."

–Light, Singer and Willett

Or, not as catchy but perhaps more accurate:

Fancy analysis can't fix a poorly designed study.

≈

Producing Data

The Role of Statistics in Producing and Analyzing Data

A **study** is an examination of a subject for the purpose of advancing knowledge.

Many studies require analysis of data. Data used in a study can arise in a number of ways. **Available data** are data that were obtained prior to the study for purposes other than those of the study. There is no guarantee that results based on available data are scientifically valid.

Statistically designed studies obtain data using a pre-specified plan that ensures that specific questions of interest are answered in a scientifically valid way.

The Role of Statistics in Producing and Analyzing Data

Drawing definitive conclusions from available data is often questionable. However, **exploratory analysis** of available data is a vital part of scientific and technological progress.

As the name suggests, in an exploratory analysis investigators explore the data to find patterns that might suggest specific questions about the phenomena under study.

Exploratory analysis cannot establish scientific validity, however. For this, statistically designed studies are needed. A good strategy is to use exploratory data analysis to identify specific questions that can then be answered by targeted statistically designed studies.

The Role of Statistics in Producing and Analyzing Data

For example, a manufacturing company that produces bronze bushings found that the surface roughness on the bushings was unacceptably variable.

To identify possible reasons, they did an **exploratory** analysis of historical production data (an example of available data). This analysis identified four process conditions that seemed to be strongly associated with surface roughness measurements: lathe operator, cutting speed, feed rate, and tool type.

With this information, they **designed** a study purposely varying these four conditions to see what the effects were on surface roughness.

Characteristics and Conduct of Designed Studies

The three main types of designed studies are

- Controlled Experiments
- Observational Studies
- Sampling Studies

All use **observational units**: entities on which measurements or observations can be made.

These observatonal units are often selected from a larger population of such units. When discussing their selection, we call them **sampling units**.

Characteristics and Conduct of Designed Studies

Proper selection of the sampling units is essential to the success of designed studies, because it will enable us to

- Get a sample that is very likely to be representative of the population.
- Quantify how far from the population results our sample results are likely to be.

Selecting Sampling Units

- **Target Population:** A collection of sampling units about which we want to draw conclusions
- **Frame:** A list of all sampling units in the target population
- **Sample:** A subset of the target population from which conclusions about the target population will be drawn

Characteristics and Conduct of Designed Studies

- **Sampling Design:** A pattern, arrangement or method used for selecting a sample of sampling units from the target population
- **Sampling Plan:** The operational plan, including the sampling design, for actually obtaining or accessing the sampling units for the study

Example 1

Suppose, prior to an election, we want to estimate the outcome in the "population" of WPI students. To do so, we select 20 WPI students and interview them. For this study:

- Sampling units: Individual students
- Target population: All WPI students
- Frame: Campus directory
- Sample: The 20 selected students

$\approx$

Characteristics and Conduct of Designed Studies

- Sampling Design: There are a large number of choices, as we will see below.
- Sampling Plan: Operational plan for deciding whom to interview, how to get them to do the interview, what to do if they can't be found, or won't talk, etc.

$\approx$

Reasons to Sample

- Lower Cost
- Shorter Time
- Little Loss of Precision
- Sometimes the Only Choice (e.g. Destructive Testing)

Probability Sampling Methods

- Simple Random Sampling Each possible sample has the same chance of selection. Good if units are homogeneous and easily accessed.
- Stratified Random Sampling Sampling units are divided into distinct strata, and a simple random sample taken separately in each. Good if units in each stratum are much more homogeneous than the population as a whole or if we want to include certain subgroups in the sample.
- Cluster Sampling Units in close proximity are grouped in clusters, and clusters are sampled. Good if there are large costs due to units being widely dispersed.

In addition, many samples are taken in stages, using any of the above methods at any stage. Such sampling is called **Multistage Sampling**.

One large study conducted monthly by the US Census Bureau is the Current Population Survey, which estimates such things as unemployment and schooling by taking a multistage cluster sample of 100,000 people in 60,000 households.

Example 2

Suppose you want to estimate the average amount spent by first term sophomores at WPI for textbooks, and that you can interview 10 students for your study. How would you choose the 10 students if:

$\approx$

Characteristics and Conduct of Designed Studies

(a) You believe the distribution of the amounts spent for textbooks is pretty consistent across all students. Simple random sample is best.

(b) You believe that textbook expenses for engineering students are substantially higher than for other majors. Stratify into two strata: engineers and others. Then take a stratified random sample.

(c) You want to be certain to obtain an estimate for humanities majors, as well as other majors. Stratify so that one stratum is humanities majors. Then take a stratified random sample.

Errors in Selecting Sampling Units

- Sampling Error "Error" inherent in the sampling process. Sampling error occurs normally (i.e., is not really an error), is always present and results from the fact that the sample is not the same as the population.
- Nonsampling Errors These are really errors and result from such failings as (i) being unable to sample from the entire population, (ii) being unable to get measurements from some selected units, or (iii) getting misleading or false measurements from some selected units.

One type of nonsampling error is **Selection Bias**:

In statistics, a bias is a distortion of the results of a statistical procedure.

Selection bias is a bias introduced to the results because the sampling method to at least some extent misses certain segments of the population.

Some Possible Errors in Selecting Sampling Units for the WPI Election Survey

- Sampling Error: The extent to which the sample is unrepresentative of the population, by chance alone.
- Selection Bias: Selection bias might occur if the survey were conducted in class. In that case, the survey would miss those who don't come to class. This would result in a bias if the views of those who come to class and of those who do not come to class differ.

An Extreme Example of Selection Bias

On July 25, 2017 CNN reported:

Chronic traumatic encephalopathy, known as CTE, was found in 99% of deceased NFL players' brains that were donated to scientific research, according to a study published Tuesday in the medical journal JAMA. . . . Out of 202 deceased former football players total – a combination of high school, college and professional players – CTE was neuropathologically diagnosed in 177, the study said.

Why is this an extreme example of selection bias?

Answer: Brains donated to scientific research are overwhelmingly from individuals exhibiting symptoms of brain damage.

Note:

If the sampling units are selected by some non-probability method (convenience, for example), the results of the study are, strictly speaking, only applicable to the sampling units in the study. In order to have results of the study apply to the target population, sampling units must be selected from that target population using an appropriate probability sampling design.

Characteristics and Conduct of Designed Studies

Now that we have acquired a sample, what do we do with it?

That depends on what kind of study we are conducting.

$\approx$

Recall:

Types of Designed Studies

- Controlled Experiment
- Observational Study
- Sampling study

$\approx$

Controlled Experiments

Before we can discuss controlled experiments, we need some terminology:

- Experimental Unit: A sampling unit selected for use in a controlled experiment.
- Response: A measurement or observation of interest that is made on an experimental unit.

Controlled Experiments

- Factor: Something thought to influence the response.
  - *o* Experimental Factor: A factor that is purposely varied by the experimenter.
  - *o* Nuisance Factor: A factor that cannot be controlled by the experimenter. Nuisance factors may or may not be known to the experimenter.

Controlled Experiments

- Level: A value assumed by a factor in an experiment.
- Treatments: The combinations of levels of experimental factors for which the response will be observed.

Controlled Experiments

We are now ready to define **Controlled Experiment**:

A **Controlled Experiment** *is a study in which treatments are imposed on experimental units in order to observe a response.*

$\approx$

Example 3

A printing company is having trouble with ink overflow on printed documents. After much brainstorming and discussion, they have narrowed the possible causes to two printing machine settings: the pressure plate setting and the ink flow rate setting. They design and run a controlled experiment to evaluate the effect of these settings on the finished product. They decide on three settings, low, medium and high, for the pressure plate and two settings, low and high, for the ink flow rate. The response is the improperly inked area on a test sheet. The data are (in $(\text{mm})^2$):

Controlled Experiments

Ink flow setting	Pressure plate setting		
	Low	Medium	High
Low	25.300 21.600	27.100 24.200	19.700 21.900
High	31.100 29.500	25.600 23.100	26.600 23.900

Controlled Experiments

For this experiment:

- Experimental units: paper sheets.
- Response: improperly inked area.
- Experimental factors: pressure plate setting, ink flow setting.
- Nuisance factors: variation in paper characteristics, environmental factors (temperature, humidity, etc.), ink supplier, etc.
- Factor levels: low, medium, high for pressure plate; low, high for ink flow rate.
- Treatments: pressure plate & ink flow rate combinations.

Controlled Experiments

This is a controlled experiment because treatments (pressure plate & ink flow rate combinations) are imposed on experimental units (paper sheets) in order to observe a response (improperly inked area).

$\approx$

Controlled Experiments

Here are some further quantities and concepts of importance in controlled experiments:

o **Effect:** The change in the average response between two factor levels or between two combinations of factor levels. Here are some effects from example 3:

Controlled Experiments

Ink flow setting	Pressure plate setting			Row Mean
	Low	Medium	High	
Low	25.300 21.600	27.100 24.200	19.700 21.900	
Mean	23.450	25.650	20.800	23.300
High	31.100 29.500	25.600 23.100	26.600 23.900	
Mean	30.300	24.350	25.250	26.633
Column Mean	26.875	25.000	23.025	24.967

The effect of high ink flow over low ink flow is
$26.633 - 23.300 = 3.333$.

Controlled Experiments

Ink flow setting	Pressure plate setting			Row Mean
	Low	Medium	High	
Low	25.300 21.600	27.100 24.200	19.700 21.900	
Mean	23.450	25.650	20.800	23.300
High	31.100 29.500	25.600 23.100	26.600 23.900	
Mean	30.300	24.350	25.250	26.633
Column Mean	26.875	25.000	23.025	24.967

The effect of high pressure plate setting over medium pressure plate setting is $23.025 - 25.000 = -1.975$.

Controlled Experiments

Ink flow setting	Pressure plate setting			Row Mean
	Low	Medium	High	
Low	25.300 21.600	27.100 24.200	19.700 21.900	
Mean	23.450	25.650	20.800	23.300
High	31.100 29.500	25.600 23.100	26.600 23.900	
Mean	30.300	24.350	25.250	26.633
Column Mean	26.875	25.000	23.025	24.967

The effect of low ink flow and high pressure plate settings over high ink flow and medium pressure plate settings is $20.800 - 24.350 = -3.55$.

Controlled Experiments

o **Confounding:** Two or more factors are **confounded** if it is impossible to separate their individual effects.

Here is an example of how confounding might occur in the ink flow experiment:

Suppose that the experimenters ran the trials for the low and high ink flow settings with different kinds of paper. Then paper type and ink flow would be confounded: any observed effect of ink flow could have been due to paper type instead.

≈

Randomized Controlled Experiments

By helping ensure that the experimental units receiving different treatments are similar in all other respects, random assignment of treatments to experimental units protects against unsuspected biases. Random assignment is beneficial in at least two other ways.

$\approx$

Randomized Controlled Experiments

- Random assignment of treatments to experimental units provides the foundation for **statistical inference**, on which many scientific applications of statistics depend. We will study statistical inference in chapters 5 and 6.
- The most compelling reason to conduct a controlled experiment is to establish **causality**: that is, to show that changing the treatment will cause a change in the response. Randomization "washes out" or "evens out" the effects of unsuspected nuisance factors, which is essential to establishing causality.

Controlled experiments in which treatments are assigned at random to experimental units are called **randomized controlled experiments**.

Another Example

Suppose we want to test two different teaching methods in MA2611 labs. To do so, we select two lab sections, one at 8am and the other at 3pm. We randomly assign one of the methods to be used in the 8am lab and the other method in the 3 pm lab. Later, we compare scores for the two sections.

- Is this a randomized controlled experiment? Yes.
- Is this a good design? Why or why not? It is not the best, because it confounds time of lab with treatment. Students in the 8am might differ in meaningful ways from those in the 3pm. So observed differences in scores might be confounded with the times of the labs.

≈

Assigning Treatments to Experimental Units: Two Commonly-Used Designs

- **Completely Randomized Design (CRD)** Treatments assigned to experimental units completely at random. Works well if units are homogeneous.
- **Randomized Complete Block Design (RCBD)** Experimental units grouped into blocks and treatments are assigned at random within each block. Effective if units within each block are more homogeneous than all units taken as a whole.

Two Different Designs for the Ink Overflow Experiment
Recall there are 6 treatments (2 ink flow, 3 pressure plate settings) and two sheets used for each, for a total of 12 experimental runs giving a total of 12 observations.

Suppose there are two printing machines on which the experiment is to be run.

- For a completely randomized design, assign all 12 runs to the two machines at random. So, for example, machine 1 might get both treatments at high ink flow setting and high pressure plate setting. Of course the order of the runs should also be randomized. This kind of design makes sense if the machines act very much the same.

Randomized Controlled Experiments

- For a randomized complete block design, assign one complete set of six treatments to machine 1 and the other to machine 2. The units assigned to each machine form a block, so there are two blocks. Run each set of treatments in random order. This kind of design makes sense if there are substantial differences between the machines.

$\approx$

More on Blocking

Here is another example of blocking:

25 pleasure boats around the country are available to test two types of marine paint. Make each boat a block by applying both types of paint to each. This (a) reduces boat-to-boat variation and variation due to such things as environment, and (b) makes the results of the study applicable to a wider range of environments and boat types.

Another Example, Continued

How might you modify the design of the study of two teaching methods in MA2611 labs to improve it?

Digging a Little Deeper...

You might notice parallels between two of the types of sampling we discussed earlier and the two types of random treatment assignment just presented.

Specifically, a completely random design for assigning treatments to experimental units is analogous to simple random sampling for obtaining those units from a population, while a randomized complete block design is analogous to stratified random sampling.

Digging a Little Deeper...

However, don't confuse sampling with treatment assignment. The purpose of the random sampling designs is to obtain a representative sample so that the results of any analysis are applicable to the population. The purpose of the random experimental designs is to remove bias due to nuisance factors and so enable accurate, causal conclusions.

Stratification and blocking are mechanisms to increase the efficiency and precision of the results.

Controlled Experiments

Experiments in which two treatments occur in each block are sometimes called **paired**. Pairing can occur in a number of ways.

For example, consider an experiment conducted by a pharmaceutical company to test a new medication designed to lower LDL cholesterol levels. Here are a couple of different ways to design a paired experiment to do this:

$\approx$

Controlled Experiments

- Ten subjects are recruited for the study. Their LDL levels are measured at the beginning of the study. They are then given the medication for one month and their LDL levels measured at the end of that time. Here, subjects are blocks and paired measurements are taken on each.
- Twenty subjects are recruited for the study in such a way that each subject is paired with another based on gender, age, LDL levels, and other relevant variables. One of each pair is randomly selected to receive the medication; the other does not receive it. After one month their LDL levels are measured. Here the paired subjects form the blocks.

Principles of Experimental Design

- Block What You Can.
- Randomize What You Cannot Block.
- Replicate as Time and Budget Permit.
  Replication=repetition. Beware of duplication.
- Confirm the Results.

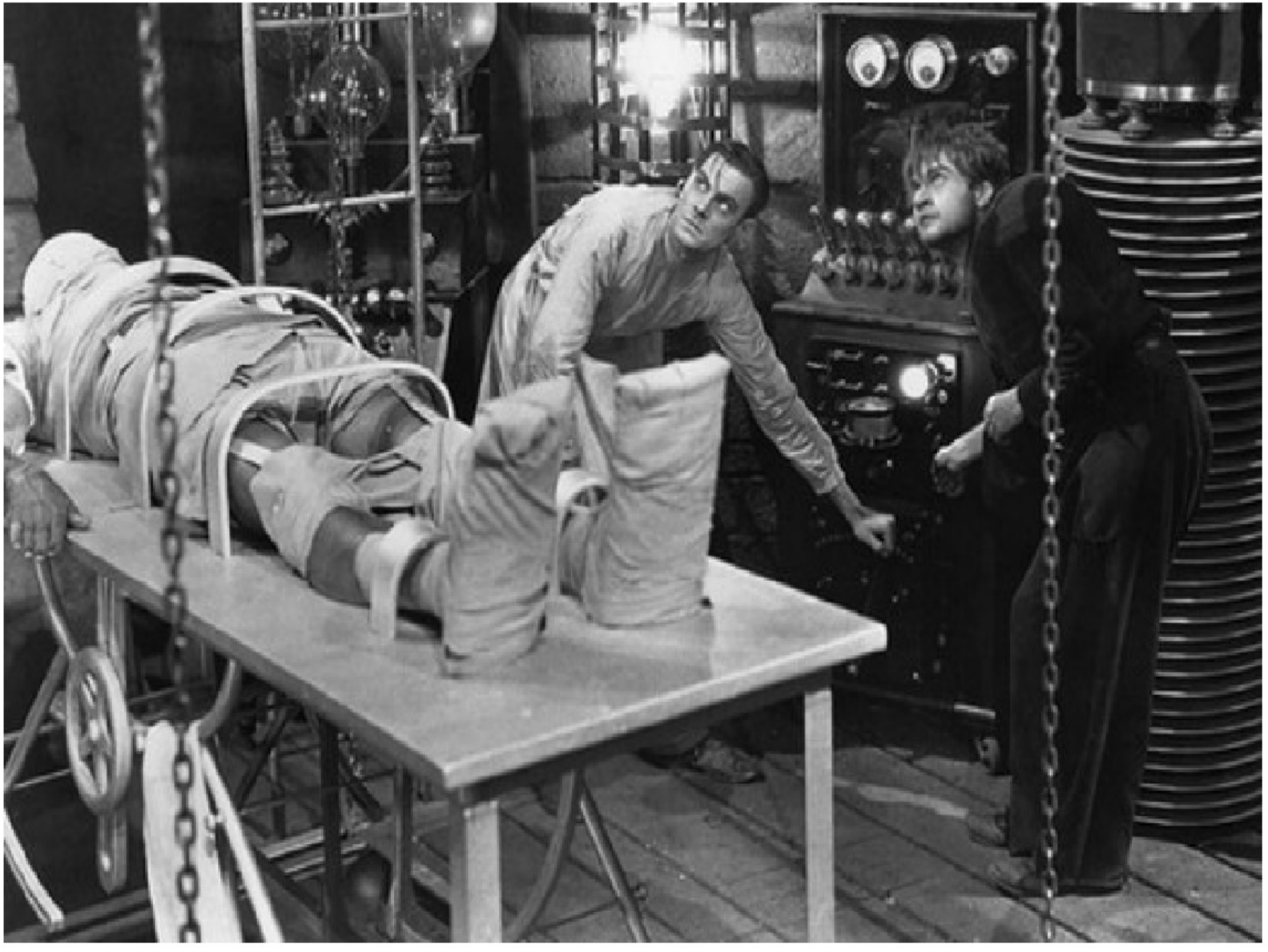
Principles of Experimental Design in the Ink Overflow Experiment

- Block What You Can. The experiment was run as an RCBD with the units assigned to the two different machines as blocks and each treatment run on each machine.
- Randomize What You Cannot Block. Paper sheets were randomly assigned and order of run was randomized in each machine (block) separately.

Controlled Experiments

- Replicate as Time and Budget Permit. Only 1 rep was done (1 observation for each treatment and each machine) due to the limited availability of the machines.
- Confirm the Results. To verify the results, confirmatory experiments were later run.

Experimenting With Human Subjects



Experimenting with Human Subjects

Experiments using human subjects have their own terminology. Here are some of the terms:

- **Treatment Group** Group of subjects that receives a treatment.
- **Control Group** Group of subjects that receives no treatment or a neutral treatment.
- **Placebo** Neutral "treatment" given to subjects in the control group. An interesting article on placebos is found here.
- **Double-Blind** Neither subject nor evaluator(s) know which treatment (if any) was given.
- **Pairing or Matching** Subjects are matched on the basis of various nuisance variables such as age, gender, or health status.

Example: Salk Vaccine Field Trial, p. 108 of the original text, p. 92 of the pdf/local version.

Observational Studies

Observational Studies

The second type of designed study we will consider is the observational study. We will study two types:

- **Cohort Study (aka Prospective Study)**
- **Case-Referent Study (aka Retrospective Study)**

$\approx$

Example 4

An example of an observational study is a study published in the **Journal of the American Medical Association**, which assessed the health patterns of 5,000 Canadians, and found that those with the greatest folic acid intake had 68% less fatal coronary disease than those with the lowest intake.

This study is not a controlled experiment since it merely **observed** the folic acid intake and coronary death outcomes, rather than assigning a folic acid regimen to individuals.

$\approx$

Cohort (aka Prospective) Studies

Also known as quasi-experiments, because they are controlled experiment "wannabees", cohort studies do not control the assignment of treatments to experimental units (e.g., we cannot assign a human subject a certain number of cigarettes per day).

As a result, cohort studies (and, in fact, all observational studies) can only demonstrate association, not cause-effect, between treatments and responses.

Cohort Studies

In a cohort study, "treatment" and "control" groups (known as **cohorts**) are established based on the hypothesized cause. The pattern of response (i.e., the effect) is then compared for the groups.

In a study of the relation between smoking and lung cancer, the cohorts might be smokers and non-smokers, and the response might be whether or not the subject develops lung cancer.

If the percentage of smokers who develop lung cancer is substantially different than the percentage of non-smokers who develop lung cancer, we would conclude there is an association between smoking and lung cancer.

Example 4, Continued:
If the Canadian folic acid study assigned individuals to groups based on their reported folic acid intake, followed them for a period of time to observe and compare the incidence of fatal coronary disease for the two groups, then it was a cohort study.

$\approx$

Case-Referent (aka Retrospective) Studies

In a case-referent study, groups are formed based on the response (i.e., the effect) and patterns in the hypothesized causes are compared from group to group.

In a study of the relation between smoking and lung cancer, we might form two groups: those who contracted lung cancer and those who did not (getting cancer or not is the effect). We might then compare the percentages of smokers in the two groups (smoking is the presumed cause).

If the percentage of smokers among the lung cancer group is substantially different than the percentage of smokers among those who did not develop lung cancer, we would conclude there is an association between smoking and lung cancer.

Case-Referent Studies

Note that in order to perform a case-referent study, the response has to have already been measured at the time the study is done.

Case-Referent studies are particularly useful when

- The time between the hypothesized cause and observed effect is large, or
- The effect occurs rarely.

$\approx$

Example 4, Continued

If the Canadian folic acid study classified individuals into groups based on whether or not they died of coronary disease, and compared the pattern of folic acid intake for the two groups, then it was a case-referent study.

$\approx$

Caution

Some students get the mistaken idea that if the study was done in the past, or done using data taken in the past, it is case-referent. Case-referent refers only to the fact that the groups are formed based on the outcome and then differences in potential causal factors are sought. The case group is the group that gets the positive outcome (e.g., cancer) and the referent group the group that gets the negative outcome (e.g., no cancer).

While it is true that outcomes must be observed prior to analyzing a case-referent study, studies based on data taken in the past can also be cohort studies if groups are established based on characteristics observed prior to observing the outcomes.

For example, suppose the researchers conducted the Canadian folic acid study by randomly selecting 5000 patient records which included folic acid intake, and whether they suffered fatal coronary disease by the end of a 5 year period. Even though the data were taken in the past, if they formed comparison groups based on folic acid intake and compared outcomes, the study is a cohort study.

$\approx$

Cause-Effect

Only a properly-designed and conducted controlled experiment can establish a cause-effect relationship between factors and response.

The biggest difference between controlled experiments and observational studies, such as cohort and case-referent studies, that attempt to show cause-effect, is the idea of **control**: the ability of the experimenter to assign treatments to experimental units. It is the control in controlled experiments that validates cause-effect conclusions and the lack of control in observational studies that casts doubt on cause-effect conclusions.

Applicability of Study Results

If we want the results of controlled experiments, cohort studies or case-referent studies to be applicable to a larger target population, we must obtain the experimental or observational units from that population using a valid sampling method.

However, there is nothing in the conduct of such studies that requires such sampling: even if the units used in the study are not selected from the target population in an appropriate way, the results of the study can still be valid for the units in the study.

Sampling Studies

In contrast, sampling studies always rely on data obtained by sampling from a larger target population. The most familiar sampling study is the sample survey.

Sample surveys

- Use a sample of sampling units obtained from a population to obtain information about the whole population.
- Have as their primary goals description of various aspects of the population from which the sample is obtained, or comparison of subgroups from that population (not establishment of association).

Sample Surveys

The study described in Example 1, in which a sample of 20 WPI students was interviewed to estimate the outcome of an election is an example of a survey.

Another is the Current Population Survey, conducted monthly by the US Census Bureau. The CPS questions about 100,000 people in some 60,000 households nationwide, and uses the results to estimate measures of the state of the nation such as income, unemployment, and schooling.

$\approx$

Sample Surveys

Note also that sample surveys are not confined to samples from human populations.

- A survey of wildlife populations might consist of monitoring a sample of habitat parcels (perhaps using capture-recapture methods).
- Landsat surveys use satellite imagery to estimate land use characteristics worldwide.
- Many manufacturers sample incoming materials or outgoing product to assess quality.
- Election monitors take samples of election ballots to evalulate the fairness of elections.

Non-sampling Errors in Studies of Human Populations

Recall our earlier discussion of sampling error (and how it's not really an "error") and selection bias which is one variety of nonsampling error.

Unless a study obtains responses from the entire target population (in which case it is called a census), sampling error is certain to occur. Selection bias may also occur.

Non-sampling Errors in Studies of Human Populations

Studies of human populations in which individuals are asked to respond to questions, verbally or in writing, are subject to other nonsampling errors, such as

- **Nonresponse bias:** Bias due to failure to obtain responses from some subjects.
- **Response bias:** Bias due to erroneous responses from some subjects.

Example 4, Continued

Here are some possible non-sampling errors in the Canadian Folic Acid Study:

- **Nonresponse bias** might occur if certain individuals refuse to supply their medical histories.
- **Response bias** might occur if the subject doesn't tell the truth about his or her health history because of concerns that giving information about poor health will adversely affect future insurability.

Recap:

- The role of statistics in producing and analyzing data
- Selecting sampling units
  - Sampling designs
  - Sampling errors
  - Designing sampling plans

- Controlled experiments
  - Principles of experimental design
  - Experimenting with human subjects
- Observational studies
  - Cohort studies
  - Case-referent studies
- Sampling studies
- Cause and effect in designed studies
- Non-sampling Errors in Studies of Human Populations