Chapter Two
Biostatistical Design of Medical Studies

2.2 Problems to be investigated

Biomedical studies arise in many ways.

- A particular study may result from a sequence of experiments.
- Observational Study.
- The study may be investigated by a governmental agency in response to a question of national importance.
- Many of the critical studies and experiments in biomedical science have come from one person with an idea for a radical interpretation of past data.

Statistical Considerations

- Statistical considerations may suggest that an experiment is too expensive to conduct, or may suggest an approach that differs from that planned.
- The need to evaluate data from a study statistically forces an investigator to sharpen the focus of the study.
- It makes one translate intuitive ideas into an analytical model capable of generating data that may be evaluated statistically.

Many different studies

- To answer a given scientific question, many different studies may be considered. Possible studies may range from small laboratory experiments, to large and expensive experiments involving humans, to observational studies.

- In laboratory research, many different experiments may shed light on a given hypothesis or question. Sometimes, less-than-optimal execution of a well-conceived experiment sheds more light than arduous and excellent experimentation unimaginatively designed.

2.3 Various Types of Studies

A problem may be investigated in a variety of ways. To decide on your method of approach, it is necessary to understand the types of studies that might be done. To facilitate the discussion of design, we introduce definitions of commonly used types of studies.

Definition 2.1. An *observational study* collects data from an existing situation. The data collection does not intentionally interfere with the running of the system.

Definition 2.2. An *experiment* is a study in which an investigator deliberately sets one or more factors to a specific level.

Experiments lead to stronger scientific inferences than do observational studies.

Definition 2.3. A *laboratory experiment* is an experiment that takes place in an environment (called a *laboratory*) where experimental manipulation is facilitated.

Definition 2.4. A *comparative experiment* is an experiment that compares two or more techniques, treatments, or levels of a variable.

Definition 2.5. An *experimental unit* or *study unit* is the smallest unit on which an experiment or study is performed.

Definition 2.6. An experiment is a *crossover experiment* if the same experimental unit receives more than one treatment or is investigated under more than one condition of the experiment. The different treatments are given during nonoverlapping time periods.

Definition 2.7. A *clinical study* is one that takes place in the setting of clinical medicine.

A study that takes place in an organizational unit dispensing health care—such as a hospital, psychiatric clinic, well-child clinic, or group practice clinic—is a clinical study.

The concepts of **prospective studies and retrospective studies**, usually involving human populations.

Definition 2.8. A *cohort* of people is a group of people whose membership is clearly defined.

Example: All persons enrolling in the Graduate School at FIU for the fall quarter of 2014.

Definition 2.9. An *endpoint* is a clearly defined outcome or event associated with an experimental or study unit.

Definition 2.10. A *prospective study* is one in which a cohort of people is followed for the occurrence or nonoccurrence of specified endpoints or events or measurements.

Definition 2.11. *Baseline characteristics* or *baseline variables* are values collected at the time of entry into the study.

Definition 2.12. A *retrospective study* is one in which people having a particular outcome or endpoint are identified and studied.

Definition 2.13. A *case–control study* selects all cases, usually of a disease, that meet fixed criteria. A group, called *controls*, that serve as a comparison for the cases is also selected. The cases and controls are compared with respect to various characteristics.

Definition 2.14. In a *matched case–control study*, controls are selected to match characteristics of individual cases. The cases and control(s) are associated with each other.

Definition 2.15. In a *frequency-matched case–control study*, controls are selected to match characteristics of the entire case sample (e.g., age, gender, year of event). The cases and controls are not otherwise associated. There may be more than one control for each case.

Definition 2.16. A *longitudinal study* collects information on study units over a specified time period. A *cross-sectional study* collects data on study units at a fixed time.

Example: Suppose that we want to study characteristics of cases of a disease. One way to do this would be to identify new cases appearing during some time interval. A second possibility would be to identify all known cases at some fixed time. The first approach is *longitudinal* ; the second approach is *cross-sectional*.

Definition 2.17. A *placebo treatment* is designed to appear exactly like a comparison treatment but to be devoid of the active part of the treatment.

Definition 2.18. The *placebo effect* results from the belief that one has been treated rather than having experienced actual changes due to physical, physiological, and chemical activities of a treatment.

Definition 2.19. A study is *single blind* if subjects being treated are unaware of which treatment (including any control) they are receiving. A study is *double blind* if it is single blind and the people who are evaluating the outcome variables are also unaware of which treatment the subjects are receiving.

2.4 Steps Necessary to perform a study

In this section we outline briefly the steps involved in conducting a study.

1. A question or problem area of interest is considered. This does not involve biostatistics per se.
2. A study is to be designed to answer the question. The design of the study must consider at least the following elements:

a. Identify the data to be collected. This includes the variables to be measured as well as the number of experimental units, that is, the size of the study or experiment.
b. An appropriate analytical model needs to be developed for describing and processing data.
c. What inferences does one hope to make from the study? What conclusions might one draw from the study? To what population(s) is the conclusion applicable?

3. The study is carried out and the data are collected.
4. The data are analyzed and conclusions and inferences about the population characteristics are drawn.
5. The results are used. This may involve changing operating procedures, publishing results, or planning a subsequent study.

2.5 Ethics

Many studies and experiments in the biomedical field involve animal and/or human participants. Moral and legal issues are involved in both areas. Ethics must be of primary concern. In particular, we mention five points relevant to experimentation with humans:

1. It is our opinion that all investigators involved in a study are responsible for the conduct of an ethical study to the extent that they may be expected to know what is involved in the study.

2. Investigators are close to a study and often excited about its potential benefits and advances. It is difficult for them to consider all ethical issues objectively. For this reason, in proposed studies involving humans (or

animals), there should be review by people not concerned or connected with the study or the investigators.

3. People participating in an experiment should understand and sign an informed consent form. The *principle of informed consent* says that a participant should know about the conduct of a study and about any possible harm and/or benefits that may result from participation in the study.

4. Subjects should be free to withdraw at any time, or to refuse initial participation, without being penalized or jeopardized with respect to current and future care and activities.

5. When possible, animal studies be done prior to human experimentation.

2.6 Data Collection: Design of Forms

2.6.1 What Data Are to Be Collected?

In studies involving only one or two investigators, there is often almost complete agreement as to what data are to be collected. In this case it is very important that good laboratory records be maintained. The necessity for keeping detailed notes is even more crucial in large studies or experiments involving many investigators; it is difficult for one person to have complete knowledge of a study.

In a large collaborative study involving a human population, it is not always easy to decide what data to collect. For example, often there is interest in getting prognostic information. How many potentially prognostic variables should you record?

Suppose that you are measuring pain relief or quality of life; how many questions do you need to characterize these abstract ideas reasonably? In looking for complications of drugs, should you instruct investigators to enter all complications?

There are many examples where too few data were collected. One of the most difficult tasks in designing forms is to remember to include all necessary items.

To assure that all necessary data are on the form, you are advised to follow four steps:

1. Perform a thorough review of all forms *with a written response* by all participating investigators.

2. Decide on the statistical analyses beforehand. Check that *specific* analyses involving *specific* variables can be run. Often, the analysis is changed during processing of the data or in the course of "interactive" data analysis.

3. Look at other studies and papers in the area being studied. It may be useful to mimic analyses in the most outstanding of these papers. If they contain variables not recorded in the new study, find out why. The usual reason for excluding variables is that they are not needed to answer the problems addressed.

4. If the study includes a pilot phase, analyze the data of the pilot phase to see if you can answer the questions of interest when more data become available.

2.6.2 Clarity of Questions

The task of designing clear and unambiguous questions is much greater than is generally realized. The following points are of help in designing such questions:

1. Who is filling out the forms? Forms to be filled out by many people should, as much as possible, be self-explanatory.

2. The degree of accuracy and the units required should be specified where possible. For example, data on heights should not be recorded in both inches and centimeters in the same place.

3. A response should be required on all sections of a form. Then if a portion of the form has no response, this would indicate that the answer was missing.

4. There are many alternatives when collecting data about humans: forms filled out by a subject, an in-person interview by a trained interviewer, a telephone interview, forms filled out by medical personnel after a general discussion with the subject, or forms filled out by direct observation.

2.6.3 Pretesting of Forms and Pilot Studies

If it is extremely difficult, indeed almost impossible, to design a satisfactory form, how is one to proceed? It is necessary to have a pretest of the forms, except in the simplest of experiments and studies. In a *pretest*, forms are filled out by one or more people prior to beginning an actual study and data collection.

A more complete approach is to have a *pilot study*, which consists of going through the actual mechanics of a proposed study. Thus, a pilot study works out both the "bugs" from forms used in data collection and operational problems within the study.

To evaluate the extent to which the data collected are understood, it is good procedure to ask others to examine some of the same study units and to record their opinion without first discussing what is meant by the categories being recorded. If there is great variability, this should lead to a need for appropriate caution in the interpretation of the data.

2.6.4 Layout and Appearance

The physical appearance of forms is important if many people are to fill them out. People attach more importance to a printed page than to a mimeographed page, even though the layout is the same. If one is depending on voluntary reporting of data, it may be worthwhile to spend a bit more to have forms printed in several colors with an attractive logo and appearance.

2.7 Data Editing and Verification

If a study involves many people filling out forms, it will be necessary to have a manual and/or computer review of the content of the forms before beginning analysis. Among checks that go into data editing are the following:

1. *Validity checks*. Check that only allowable values or codes are given for answers to the questions. For example, a negative weight is not allowed. A simple extension of this idea is to require that most of the data fall within a given range; range checks are set so that a small fraction of the valid data will be outside the range and will be "flagged.

2. *Consistency checks*. There should be internal consistency of the data. Following are some examples:

a. If more than one form is involved, the dates on these forms should be consistent with each other (e.g., a date of surgery should precede the date of discharge for that surgery).
b. Consistency checks can be built into the study by collecting crucial data in two different ways (e.g., ask for both date of birth and age).
c. If the data are collected sequentially, it is useful to examine unexpected changes between forms (e.g., changes in height, or drastic changes such as changes of weight by 70%). Occasionally, such changes are correct, but they should be investigated.
d. In some cases there are certain combinations of replies that are mutually inconsistent; checks for these should be incorporated into the editing and verification procedures.

3. *Missing forms*. In some case–control studies, a particular control may refuse to participate in a study. Some preliminary data on this control may already have been collected. Some mechanism should be set up so that it is clear that no further information will be obtained for that control.

2.8 Data Handling

All except the smallest experiments involve data that are eventually processed or analyzed by computer. Forms should be designed with this fact in mind. It should be easy to enter the form by keyboard. Some forms are called *self-coding*: Columns are given next to each variable for data entry.

For very large studies, the logistics of collecting data, putting the data on a computer system, and linking records may hinder a study more than any other factor. In any large study, people with expertise in data handling and computer management of data should be consulted during the design phase. Inappropriately constructed data files result in unnecessary expense and delay during the analytic phase.

Computer files or tapes will occasionally be erased accidentally. In the event of such a disaster it is necessary to have backup computer tapes and documentation.

Data collection and handling usually involves almost all participants of the study and should not be underestimated. It is difficult to give a rule of thumb, but in a wide variety of studies, 15% of the expense has been in data handling, processing, and analysis.

2.9 Amount of Data Collected: Sample Size

One of the tasks of a statistician is to determine an appropriate sample size for a study. If the purpose of an experiment is to estimate some quantity, there is a need to know how precise an estimate is desired and how confident the investigator wishes to be that the estimate is within a specified degree of precision. If the purpose of an experiment is to compare several treatments, it is necessary to know what difference is considered important and how certain the investigator wishes to be of detecting such a difference. Statistical calculation of sample size requires that all these considerations be quantified.

2.10 Inferences from a Study

2.10.1 Bias

The statistical term *bias* refers to a situation in which the statistical method used does not estimate the quantity thought to be estimated or test the hypothesis thought to be tested.

Consider some examples of biased statistical procedures:

1. A proposal is made to measure the average amount of health care in the United States by means of a personal health questionnaire that is to be passed out at an American Medical Association convention. In this case, the AMA respondents constitute a biased sample of the overall population.

2. A famous historical example involves a telephone poll made during the Dewey–Truman presidential contest. At that time—and to some extent today—a large section of the population could not afford a telephone. Consequently, the poll was conducted among more well-to-do citizens, who constituted a biased sample with respect to presidential preference.

2.10.2 Similarity in a Comparative Study

In a comparative experiment, we would like to try out experiments on similar units. We now discuss similarity where it is assumed for the sake of discussion that the experimental units are humans. The ideas and results, however, can be extended to animals and other types of experimental units.

The experimental situations being compared will be called *treatments*.
To get a fair comparison, it is necessary that the treatments be given to similar units.

Of all human beings, *identical twins* are the most alike, by having identical genetic background. Often, they are raised together, so they share the same

environment. In an observational twin study, a strong scientific inference can be made if enough appropriate pairs of identical twins can be found.

A second approach is that of matching or pairing individuals. The rationale behind *matched* or *matched pair studies* is to find two persons who are identical with regard to all "pertinent" variables under consideration except the treatment.

A third approach is not to match on specific variables but to try to select the subjects on an intuitive basis. For example, such procedures often select the next person entering the clinic, or have the patient select a friend of the same gender.

Still another approach, even farther removed from the "identical twins" approach, is to select a group receiving a given treatment and then to select in its entirety a second group as a control.

The final approach is to select the two groups in some manner realizing that they will not be similar, and to measure pertinent variables, such as the variables that one had considered matching upon, as well as the appropriate endpoint variables.

None of the foregoing methods of obtaining "valid" comparisons are totally satisfactory. In the 1920s, Sir Ronald A. Fisher and others made one of the great advances in scientific methodology—they assigned treatments to patients by chance; that is, they assigned treatments *randomly*. The technique is called *randomization*.

2.10.3 Inference to a Larger Population

Usually, it is desired to apply the results of a study to a population beyond the experimental units. To extend results to a larger population, experimental units should be *representative* of the larger population.

Sometimes, the results of a technique are compared with "historical" controls; that is, a new treatment is compared with the results of previous patients using an older technique.

The results of an observational study carried out in one country may be extended to other countries. This is not always appropriate. Much of the "bread and butter" of epidemiology consists of noting that the same risk factor seems to produce different results in different populations.

2.10.4 Precision and Validity of Measurements

Statistical theory leads to the examination of variation in a method of measurement. The variation may be estimated by making repeated measurements on the same experimental unit. If instrumentation is involved, multiple measurements may be taken using more than one of the instruments to note the variation between instruments. If different observers take measurements, a quantification of the variability between observers may be made.

Statistics helps in thinking about alternative methods of measuring a quantity. When introducing a new apparatus or new technique to measure a quantity of interest, validation against the old method is useful. In considering subjective ratings by different people, it often turns out that a quantity is not measured in the same fashion if the measurement method is changed.

2.10.5 Quantification and Reduction of Uncertainty

Because of variability, there is uncertainty associated with the interpretation of study results. Statistical theory allows quantification of the uncertainty. If a quantity is being estimated, the amount of uncertainty in the estimate must be assessed. In considering a hypothesis, one may give numerical assessment of the chance of occurrence of the results observed when the hypothesis is true.

Appreciation of statistical methodology often leads to the design of a study with increased precision and consequently, a smaller sample size. An example of an efficient technique is the statistical idea of blocking. Blocks are subsets of relatively homogeneous experimental units. The strategy is to apply all treatments randomly to the units within a particular block. Such a design is called a *randomized block design*.