CHAPTER ONE

GENERAL OVERVIEW

Statistics is the science whereby inferences are made about specific random phenomena on the basis of relatively limited sample material. The field of statistics can be subdivided into two main areas: mathematical statistics and applied statistics. **Mathematical statistics** concerns the development of new methods of statistical inference and requires detailed knowledge of abstract mathematics for its implementation. **Applied statistics** concerns the application of the methods of mathematical statistics to specific subject areas, such as economics, psychology, and public health. **Biostatistics** is the branch of applied statistics that concerns the application of statistical methods to medical and biological problems.

A good way to learn about biostatistics and its role in the research process is to follow the flow of a research study from its inception at the planning stage to its completion, which usually occurs when a manuscript reporting the results of the study is published. As an example, I will describe one such study in which I participated.

A friend called one morning and in the course of our conversation mentioned that he had recently used a new, automated blood-pressure device of the type seen in many banks, hotels, and department stores. The machine had read his average diastolic blood pressure on several occasions as 115 mm Hg; the highest reading was 130 mm Hg. I was horrified to hear of his experience, since if these readings were true, my friend might be in imminent danger of having a stroke or developing some other serious cardiovascular disease. I referred him to a clinical colleague of mine who, using a standard blood-pressure cuff, measured my friend's diastolic blood pressure as 90 mm Hg. The contrast in the readings aroused my interest, and I began to jot down the readings on the digital display every time I passed the machine at my local bank. I got the distinct impression that a large percentage of the reported readings were in the hypertensive range. Although one would expect that hypertensives would be more likely to use such a machine, I still believed that blood-pressure readings obtained with the machine might not be comparable with those obtained using standard methods of blood-pressure measurement. I spoke to Dr. B. Frank Polk about my suspicion and succeeded in interesting him in a small-scale evaluation of such machines. We decided to send a human observer who was well trained in blood-pressure measurement techniques to several of these machines. He would offer to pay subjects 50¢ for the cost of using the machine if they would agree to fill out a short questionnaire and have their blood pressure measured by both a human observer and the machine.

At this stage we had to make several important decisions, each of which would prove vital to the success of the study. The decisions were based on the following questions:

(1) How many machines should we test?

(2) How many people should we test at each machine?

(3) In what order should the measurements be taken—should the human observer or the machine be used first? Ideally, we would have preferred to avoid this problem by taking both the human and machine readings simultaneously, but this procedure was logistically impossible.

(4) What other data should we collect on the questionnaire that might influence the comparison between methods?

(5) How should the data be recorded to facilitate their computerization at a later date?

(6) How should the accuracy of the computerized data be checked?

We resolved these problems as follows:

(1) and (2) We decided to test more than one machine (four to be exact), since we were not sure if the machines were comparable in quality. However, we wanted to sample enough subjects from each machine so that we would have an accurate comparison of the standard and automated methods for each machine. We tried to predict how large a discrepancy there might be between the two methods. Using the methods of sample-size determination discussed in this book, we calculated that we would need 100 subjects at each site to have an accurate comparison.

(3) We then had to decide in what order the measurements should be taken for each person. According to some reports, one problem that occurs with repeated blood-pressure measurements is that people tense up at the initial measurement, yielding higher blood pressure than at subsequent repeated measurements. Thus, we would not always want to use the automated or manual method first, since the effect of the method would get confused with the order-of-measurement effect. A conventional technique that we used here was to **randomize** the order in which the measurements were taken, so that for any person it was equally likely that the machine or the human observer would take the first measurement. This random pattern could be implemented by flipping a coin or, more likely, by using a table of **random numbers** as appears in Table 4 of the Appendix.

(4) We felt that the major extraneous factor that might influence the results would be body size, since we might have more difficulty getting accurate readings from people with fatter arms than from those with leaner arms. We also wanted to get some idea of the type of people who use these machines; so we asked questions about age, sex, and previous hypertensive history.

(5) To record the data, we developed a coding form that could be filled out on site and from which data could be easily entered on a computer terminal for subsequent analysis. Each person in the study was assigned an identification (ID) number by which the computer could uniquely identify that person. The data on the coding forms were then keyed and verified. That is, the same form was entered twice, and a comparison was made between the two records to make sure they were the same. If the records were not the same, the form was reentered.

(6) After data entry we ran some editing programs to ensure that the data were accurate. Checking each item on each form was impossible because of the large amount of data. Alternatively, we checked that the values for individual variables were within specified ranges and printed out aberrant values for manual checking. For example, we checked that all blood-pressure readings were at least 50 and no more than 300 and printed out all readings that fell outside this range.

After completing the data-collection, data-entry, and data-editing phases, we were ready to look at the results of the study. The first step in this process is to get a general feel for the data by summarizing the information in the form of several descriptive statistics. This descriptive material can be numerical or graphical. If numerical, it can be in the form of a few summary statistics, which can be presented in tabular form or, alternatively, in the form of a **frequency distribution**, which lists each value in the data and how frequently it occurs. If graphical, the data are summarized pictorially and can be presented in one or more figures. The appropriate type of descriptive material will vary with the type of distribution considered. If the distribution is **continuous**, that is, if there are essentially an infinite number of possible values, as would be the case for blood pressure, then means and standard deviations might be the appropriate descriptive statistics. However, if the distribution is **discrete**, that is, if there are only a few possible values, as would be the case for sex, then percentages of people taking on each value would be the appropriate descriptive measure. In some cases both types of descriptive statistics are used for continuous distributions by condensing the range of possible values into a few groups and giving the percentage of people that fall into each group (e.g., the percentages of people that have blood pressures between 120 and 129 mm Hg and between 130 and 139 mm Hg, etc.).

In this study we decided first to look at mean blood pressure for each method at each of the four sites. Table 1.1 summarizes this information [1].

TABLE 1.1 Mean blood pressures and differences between machine and human readings at four locations

		Systolic blood pressure (mm Hg)					
		Machine		Human		Difference	
Location	Number of people	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
A	98	142.5	21.0	142.0	18.1	0.5	11.2
B	84	134.1	22.5	133.6	23.2	0.5	12.1
C	98	147.9	20.3	133.9	18.3	14.0	11.7
D	62	135.4	16.7	128.5	19.0	6.9	13.6

Source: By permission of the American Heart Association, Inc.

You might notice from this table that we did not obtain meaningful data from all the 100 people interviewed at each site, since we could not obtain valid readings from the machine for many of the people. This type of missing-data problem is very common in biostatistics and should be anticipated at the planning stage when deciding on sample sizes (which was not done in this study).

Our next step in the study was to determine whether the apparent differences in blood pressure between machine and human measurements at two of the locations (C, D) were "real" in some sense or were "due to chance." This type of question falls into the area of **inferential statistics**. We realized that although there was a 14-mm Hg difference in mean systolic blood pressure between the two methods for the 98 people we interviewed at location C, this difference might not hold up if we interviewed 98 other people at a different time, and we wanted to have some idea as to the **error in the estimate** of 14 mm Hg. In statistical jargon this group of 98 people represents a **sample** from the **population** of all people who use that machine. We were interested in the population and we wished to use the sample to help us learn something about the population. In particular, we wanted to know how different the **estimated** mean difference of 14 mm Hg in our sample was likely to be from the **true** mean difference in the population of all people who might use this machine. More specifically, we wanted to know if it was still possible that there was no underlying difference between the two methods and that our results were due to chance. The 14-mm Hg difference in our group of 98 people is referred to as an **estimator** of the true mean difference (d) in the population. The problem of inferring characteristics of a population from a sample is the central concern of statistical inference and is a major topic in this text. To accomplish this aim, we needed to develop a **probability model**, which would tell us how likely it is that we would obtain a 14-mm Hg difference between the two methods in a sample of 98 people if there were no real difference between the two methods over the entire population of users of the machine. If this probability were sufficiently small, then we would begin to believe that a real difference existed between the two methods. In this particular case, using a probability model based on the t distribution, we were able to conclude that this probability was less than 1 in 1000 for each of machines C and D. This probability was sufficiently small for us to conclude that there was a real difference between the automatic and manual methods of taking blood pressure for two of the four machines tested.

We used a statistical package to perform the preceding data analyses. A package is a collection of statistical programs that describe data and perform various statistical tests on the data. Currently the most widely used statistical packages include SAS, $SPSS^X$, BMDP, and Minitab.

The final step in this study, after completing the data analysis, was to compile the results in the form of a publishable manuscript. Inevitably, because of space considerations, much of the material developed during the data-analysis phase was weeded out and only the essential items were presented for publication.

The review of this study should give you some idea of what medical research is about and what the role of biostatistics is in this process. The material in this text parallels the description of the data-analysis phase of the study described. Chapter 2 summarizes different types of descriptive statistics. In Chapters 3 through 5, some basic principles of probability and various probability models for use in later discussions of inferential statistics are presented. In Chapters 6 through 13, the major topics of inferential statistics as used in biomedical practice are discussed. Issues of study design or data collection are brought up only as they relate to other topics discussed in the text.

Reference

[1] Polk, B. F., Rosner, B., Feudo, R., & Vandenburgh, M. (1980). An evaluation of the Vita-Stat automatic blood pressure measuring device. *Hypertension*, *2*(2), 221–227.