CHAPTER

1

Reporting Statistical Data

CHAPTER OUTLINE

A. Introduction
   1. Statistics and Data
   2. Scope of Book
B. Statistical Data Terms and Definitions
   1. Population vs. Sample
   2. Constant vs. Variable
   3. Nominal vs. Ordinal Data
   4. Qualitative vs. Quantitative Variables
   5. Discrete vs. Continuous Data
   6. Ungrouped vs. Grouped Data
   7. Descriptive vs. Inferential Statistics
   8. Morbidity vs. Mortality
   9. Demographic Variables
   10. Vital Statistics
C. Computerized Data
   1. Use
   2. Accuracy
D. Patient Data Collection
   1. Types of Data Collected
E. Abbreviations
   1. Patient Care
   2. Statistical
   3. Clinical Units (Some of the More Common Designations)
   4. Non-Official Abbreviations
F. Uses of Data
G. Summary
H. Chapter 1 Test

LEARNING OBJECTIVES

After studying this chapter, the learner should be able to:

1. Define "statistics."
2. Define "data."
3. Define:
   a. Demography and demographic variables
   b. Vital statistics
4. Distinguish clearly between:
   a. Population and sample.
   b. Variable and constant.
   c. Qualitative and quantitative data.
   d. Ungrouped and grouped data.
   e. Descriptive and inferential statistics.
   f. Nominal and ordinal data.
   g. Discrete and continuous data.
   h. Morbidity and mortality.
5. Identify abbreviations used in health care statistics.
6. Describe various uses of data.

People are exposed daily to some type of statistical data or statistical terms that are gathered and reported not only by the news media but also in the job arena. This is especially the case for those who work in the health care industry, where patient care data and statistics are compiled on a daily basis. Once we understand the meaningfulness of this data, we can become better managers and collectors of the data, thereby assuring appropriate uses for information.

A. INTRODUCTION

1. Statistics and Data

Statistics: A basic definition of statistics is "the mathematics of the collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling."

Statistics is defined more broadly as a branch of applied mathematics, concerned with scientific methods for collecting, organizing, summarizing, and analyzing data. The term is frequently used to refer to recorded data, for example, reports that are issued regarding traffic accident statistics or the number of outpatients treated at an outpatient clinic. Statistics is also considered a branch of study that involves the theory, methodology, and mathematical calculation concerning the collection of various kinds of data.

Reasonable decisions and valid conclusions may be drawn based on the analysis of statistical data. Statistics therefore involves both numbers and the techniques and procedures to be followed in collecting, organizing, analyzing, interpreting, and presenting information in a numerical form.

Though the term statistics is a broad term, it is narrowed and defined by its representative data, such as accident statistics, hospital statistics, employment statistics, vital statistics, and several other descriptors.

Data: Data is defined as "information, especially information organized for analysis or used as the basis for a decision; numerical information." Data are those facts that any particular situation affords or gives to an observer. Some sources define data as raw facts and figures that are meaningless in and of themselves and refer to information as meaningful data—knowledge resulting from processing data.

The term data is generally and preferably the plural of the singular datum, though it is accepted in the singular construction as well. From this term references become more specific, for example, *data base* (also called *data bank*), which is a collection of data often arranged for ease and speed of retrieval. The preparation of information for processing by computers is referred to as *data processing*.

Enormous amounts of data and numbers are collected and tabulated daily in a hospital. A record is kept of most of the transactions that occur, including the number of patients admitted, the number of electrocardiograms performed, the number of babies born, the number of patients undergoing surgery, the number of patients who die, ad infinitum.

For this collected data to be useful and meaningful, various statistical methods and formulae must be applied.

Data are collected on inpatients, outpatients, emergency room patients, employees, and so on. Collected data must be compiled into a form that will have significance and that can be used to make comparisons for decision making.

2. Scope of Book

The purpose of this textbook is to introduce the reader to the terms, formulae, and computations used for hospital statistics, with the major emphasis on inpatient hospital statistics. Much of what applies to hospital inpatient statistics can be equally applied to outpatient data collection and statistical treatment of that data. As outpatient treatment has increased enormously during the past decade and as hospital inpatient admissions have declined, more and more data are handled daily, increasing the volume of numbers and data collected over a period of time—whether it be hourly, daily, weekly, monthly, quarterly, or yearly.

The major focus of this book is the statistical treatment of inpatient hospital statistics, with emphasis on definitions, formulae, and computations. It is to be assumed that the data referred to in this book are inpatient hospital data unless otherwise specified.

It is anticipated that the book's content and problems will be useful to hospital personnel whose function is the collection and interpretation of numerical data, especially health information personnel. Often the Health Information Department is the depository for medical information and the department is frequently responsible for compiling, collecting, and organizing data. This textbook provides material and problems to facilitate the processing and interpretation of these numerical data by the responsible personnel.

It should also be noted that those responsible for data collection should make sure to collect neither too much nor too little data. Data that are never used are not worth the added expense of collecting and processing them. In other words, cost effectiveness is achieved when the information is useful and of value to an individual or to a group.

B. STATISTICAL DATA TERMS AND DEFINITIONS

It is important to acquire a knowledge of common, universal terms and definitions which apply to an area of study. Throughout this textbook, the reader will be introduced to many terms and definitions, primarily related to the health care industry and the statistical concepts employed in health care. It is important that a term have the same meaning to all who use the term. Every area of study has its own terms whether it be the study of medicine, computers, a foreign language, or health care statistics. For effective communication it is important that all speak the same language and, to that end, the reader will be introduced to many terms throughout this text.

1. Population vs. Sample

Population: The term population refers to an entire group. A population is a set of persons (or objects) having a common observable characteristic.

Every ten years the United States Census Bureau conducts a population census. Each house and residence in the United States is sent a questionnaire to be completed and returned, indicating the number of inhabitants residing at that site. Sites failing to complete the questionnaire are visited by census takers in an attempt to get as accurate a count as possible. A hospital is also an example of a specific population—a group of people admitted for the purpose of receiving medical treatment and care. A

population may also be comprised of all patients suffering from a specific disease or undergoing a specific form of treatment, such as radiotherapy.

Sample: A sample is a subset or small part of a population. Often information obtained from a sample is used to generalize from it to the entire population. A transcription supervisor lacks the time to check the accuracy of every report transcribed by each transcriptionist. It is virtually unfeasible to check every word on every report transcribed by all transcriptionists every day. Therefore, a sample is taken from the transcribed reports—say, two reports, or 5 percent of the transcribed reports—and the accuracy and quality of the transcriptionist's work is based on this sample.

The majority of the data in this textbook will focus on population statistics, in which all the patients in a specific hospital will be referred to as the population. When handling information such as mortality (also referred to as death) statistics, census data, and pregnancy data, all cases will be included in the statistical treatment rather than every fifth case or tenth case, which makes use of sampling techniques. When employing sampling statistics, it is common to *infer* that this sample is representative of a given population (like an employee's work) and *deductions* are made relative to this sample. Probability analyses and deductive statistics will not be included in this textbook.

2. Constant vs. Variable

Constant: A constant is something that assumes only one value; it is a value which is replaceable by one and only one number.

A constant is that which does not change and has one and only one value. A constant is one's date of birth or any value or specific that applies to everyone in the distribution.

Variable: A variable is something that can change, in contrast to a constant, which remains the same.

Variables are often expressed as symbols, such as X, x, Y, y, N, which can be replaced by a single number from a set of applicable numbers. Often it becomes desirable to compare variables and determine the relationship between them. For example, it may be useful to compare one variable, such as age, with another variable, such as occupation, or severity of illness, or a specific diagnosis.

3. Nominal vs. Ordinal Data

Nominal Data: The term nominal pertains to "name." Whatever distinguishing symbols are used to define a group or an individual is nominal data. These symbols often are numbers, though they can be words, designs or pictures as well. In the age of computers people constantly acquire new numbers that distinguish them from others. Examples of these distinguishing numbers are telephone numbers, zip code, social security number, driver's license number, and credit card numbers. None of these numbers represent an amount or quantity. Such numbers are used as identifiers and are referred to as nominal numbers. It is inappropriate to perform arithmetic operations on nominal data.

Ordinal: Ordinal refers to "order" or "rank." An ordinal number represents a specified (or ordered) position in a numbered series, such as an ordinal rank of seven. If it is stated that cancer is the third leading cause of death in the United States, three is the ordinal number. Some competitive events are judged based on certain criteria (div-

ing, band competition, figure skating) in which the contestant(s) is rated and scored based on rank. Grouping into low, middle, or high scores involves the ordinal scale.

4. Qualitative vs. Quantitative Variables

Qualitative Variables: Qualitative variables yield observations that can be categorized according to some characteristic or quality. Examples of this type of variable include a person's occupation, marital status, education level, race, etc.

Quantitative Variables: Quantitative variables yield observations that can be measured. Examples of this type of variable are height, weight, blood pressure, serum cholesterol, heart rate, etc. Quantitative data can be subdivided into discrete and continuous data.

5. Discrete vs. Continuous Data

Discrete Data: Discrete data are always expressed as a whole number or integer. Discrete data are most commonly obtained by counting—the number of teeth in the mouth, the number of keratoses on the skin, the number of shares traded on the New York Stock Exchange. If the variable is fixed by counting essentially indivisible units, the variable is discrete. In other words, it is a number without a fractional or decimal subdivision.

Continuous Data: Continuous variables are those that fall into the category of "measured to the nearest." The underlying scale by which measurement can be subdivided could go on indefinitely, but most data are only subdivided to a designated degree. For example, if someone were asked to measure the distance from home to work, the distance could be recorded differently, depending on the specificity required. To illustrate, the distance to the nearest mile is two miles; to the nearest half mile, 2½ miles; to the nearest quarter mile, 2¼ miles; to the nearest eighth of a mile, 2⅜ mile. Data measured in decimal fractions, but recorded to the nearest whole number, are still continuous data. Height, weight, and age are all continuous variables. A person two months away from their 22nd birthday is actually closer to age 22 than to age 21, but in most instances that person would be considered to be age 21 until their actual 22nd birthday. An individual whose height measures 5 feet 4¾ inches is closer to being 5′5″ than 5′4″.

6. Ungrouped vs. Grouped Data

Ungrouped Data: Ungrouped data is a listing of all scores as they are obtained. Ungrouped data also refers to a distribution in which scores are ranked from highest to lowest or lowest to highest but each score has its own place in the array.

Grouped Data: Grouped data involves some type of grouping or combining of scores. The most common means of grouping is the counting or tallying of like scores. In this method, all identical scores are tallied and the number recorded after the score. If five pediatric patients were all admitted on the same day and two were 10 years of age, then two tally marks would be placed in the 10-year-old age column.

With a large range of scores, it often becomes necessary to combine some scores together and reduce the spread. Ages, even when recorded to the nearest whole number, would range from newborn to over 100 years of age. With a large number of scores, it becomes necessary to group and tally scores and thus narrow the range.

Ages are often grouped, and may include a range by decade or some other grouping, say, newborn to 4 years; 5 years to 13 years; 14 to 21; 22 to 34; 35 to 49; 50 to 64; 65 to 79; 80 to 100.

7. Descriptive vs. Inferential Statistics

Descriptive Statistics: Descriptive statistics describe and analyze a given group without drawing any conclusions or inferences about a larger group. Once data has been assembled and tabulated according to some useful categories, it then needs to be summarized to determine the general trend of the data. Descriptive statistics deal with data that are enumerated, organized, and possibly graphically represented. The decennial census carried out by the United States government is an example of descriptive statistics. That data gathered are obtained and then compiled into some type of table or graph.

Inferential Statistics: Inferential statistics give information regarding kinds of claims or statements that can be reasonably made about the population based on data from a sample. Inferential statistics are concerned with reaching conclusions. At times the information available is incomplete and generalizations are reached based on the data available. When generalizations about a population are made based on information obtained from a sample, inferential statistics are utilized. A common example relates to inferences about a population based on opinion polls. This type of statistical treatment is most frequently found in more advanced statistical texts.

8. Morbidity vs. Mortality

Morbidity: Morbidity data refers to disease statistics and is gathered to provide data on the prevalence of disease. Morbidity data is far more difficult to gather than mortality (death) data due to the lack of an adequate universal state and national reporting system. Additional information regarding morbidity data gathering is provided in the chapter which includes Vital Statistics.

Mortality: Mortality refers to death statistics. The death certificate identifies the state in which the death occurred and the date of death. An entire chapter is devoted to computation of death rates and additional information on death certificates is provided in the section on Vital Statistics in a future chapter.

9. Demographic Variables

Demography is the study of characteristics of human populations. Demographic variables include the size of a population and how it changes over time; the composition of the population such as the age, sex, ethnicity, income, and health status of its members; and geographic density. As inner city residents became more affluent, families fled the inner city and moved to the suburbs, leaving the less affluent behind. This emigration to the suburbs changed the demographics of the city. Demographic data are invaluable in program planning and disease control. Demographic data are also invaluable to hospital administrators in their attempt to provide the services most needed in their communities and the areas they serve.

10. Vital Statistics

Vital statistics refers to data that records significant events and dates in human life. This data includes births, deaths, marriages and divorces. Measures of illness and dis-

ease (morbidity) also fall under the umbrella term, vital statistics. A more detailed analysis and reporting of vital statistics information is provided in future chapters.

C. COMPUTERIZED DATA

1. Use

More and more data collections and computations are being carried out by computers, using both personal computers and on-line computers connected to a central mainframe. Local area networks (LANs) are increasingly being installed. As the size of a health care facility increases, the amount of data collected also increases and this collection is facilitated by computers. Even smaller institutions are finding it profitable to invest in computers that can be accessed at any time to print out the latest statistical information, such as the census, percentage of occupancy, and other facts that management needs for decision making.

2. Accuracy

Accuracy is important when entering data either manually or by computer. Quality control measures should be incorporated to maintain correct data entry and accuracy. One should always ask whether the resultant figure from any computation is plausible and, if not, recheck the data entries.

D. PATIENT DATA COLLECTION

1. Types of Data Collected

Computerization in health care facilities has increased dramatically during the past decade and this trend will continue well into the future, making it easier to collect more data. The increased amount of information can be useful in decision making. The types of patient data that are collected in health care facilities can be classified into six broad categories, as follows:

a. Dates

Examples of dates included in this category are the patient's date of birth, date of admission, date of discharge, date of a surgical procedure, dates of various forms of treatment (both inpatient and outpatient), and date of delivery (giving birth).

b. Counts

Examples of counts include the number of patients admitted on a certain date or discharged on a certain date, the number of CBCs (complete blood counts) performed or EKGs (electrocardiograms) or any number of other tests, the number of patients receiving physical therapy treatment or chemotherapy, the number of babies delivered live or aborted, the number of patients who died in the hospital or were treated in the emergency room.

c. Test Results

Laboratory tests are a major data collection component of inpatient and outpatient examinations. These include hematology tests such as CBC, WBC (white blood

cell) differential, and RBC (red blood cell) morphology; blood chemistries such as blood glucose, BUN (blood urea nitrogen), and alkaline phosphatase; UA (urinalysis); CSF (cerebrospinal fluid) analysis; bone marrow tests; blood typing, serology, toxicology, and many more.

d. Diagnoses

Patients upon admission are assigned an admitting diagnosis (also called provisional or tentative diagnosis). Discharge diagnoses are assigned at the time of discharge and include the principal diagnosis and other diagnoses and complications. Each consultant who sees the patient provides diagnoses for their specialty area. Surgeons assign preoperative and postoperative diagnoses at the time of surgery. Diagnoses are assigned code numbers from which a disease and procedure index/data base are generated. Counts can be made for a specific disease to ascertain how many patients were diagnosed with that disorder in the period specified.

e. Procedures

If a patient undergoes a surgical procedure or diagnostic procedure, it is recorded, and most of these procedures are assigned code numbers as well. Totals can be generated for specific procedures (such as gastroscopies, mammographies, and hysterectomies) in a manner similar to that used for diagnoses.

f. Treatment Outcomes and Assessments

Upon discharge, a note is often written on a patient's medical record about the condition of the patient at the time of discharge and whether the patient was discharged home in good condition, transferred to another facility (nursing home, another hospital), or expired. Results of treatment can be recorded and various modalities of treatment can be compared based on these data. Treatment outcomes of one institution can also be compared with those of another and serve as the basis for research studies.

E. ABBREVIATIONS

Certain abbreviations are routinely used by hospitals with regard to data collection and analysis. Listed below, for easy reference, are some common abbreviations used throughout this text.

1. Patient Care

AMA	against medical advice	(patient left without a discharge order)
DOA	dead on arrival	
ER	emergency room	
IP	inpatient	
NB	newborn	
OB	obstetrical	
OP	outpatient	

2. Statistical

ADM	admission	(patient admitted to the hospital)
DIS or DC	discharge	(patient discharged from the hospital)
A&D	admitted and discharged	(patient was admitted and discharged on the same day)
	Also called I&O *(in and out) in some facilities; others refer to such patients as "come and go." In this text they will be designated as A&D.*	
A&C	*adults and children*	
	This designation is used to refer to all patients other than newborns. It is used to separate patients into two categories—newborns and others. This designation is needed because many formulae require separate computations for the two groups—newborns vs. all other patients (A&Cs). The two populations have unique characteristics and need to be treated separately.	
TRF-in	transferred in	(patient transferred into a clinical unit)
TRF-out	transferred out	(patient transferred out of a clinical unit)
>	greater than	
<	less than	
c̄	with	(from the Latin word *cum*, meaning "with")
s̄	without	(from the Latin word *sine*, meaning "without")
Σ	summation	(The uppercase Greek letter sigma means summation—it indicates that whatever follows the sign is to be added.)

3. Clinical Units (Some of the More Common Designations)

CCU	coronary care unit	OPHTH	ophthalmology
ENT	ear-nose-throat	ORTHO	orthopedics
GYN	gynecology	PED	pediatrics
ICU	intensive care unit	PSYCH	psychiatry
MED	medical care unit	REHAB	rehabilitation
NEURO	neurology/neurosurgery	SURG	surgical care unit
OB	obstetrics	UROL	urology
ONCO	oncology		

4. Non-Official Abbreviations

Throughout this text there will be abbreviations used which may not be used in all health care facilities but which facilitate computations that will be carried out in the various chapters of the text. Rather than stating the same words over and over, using an abbreviation facilitates brevity (or conciseness). Complete explanations describing each of these terms will be included in the chapters in which they are used. They are

listed here for easy reference. For the sake of brevity, the following abbreviations will be used:

Cor	coroner/medical examiner case
CTT	census-taking time
DD	discharge days
DIPC	daily inpatient census
HP	hospital pathologist
IPSD	inpatient service day
LOS	length of stay

F. USES OF DATA

Data are used in a variety of ways, for example, to justify the opening or closing of clinical units in a hospital and to assess and justify the need for new equipment, facilities, and staff. Data are invaluable to physicians in determining the proper diagnosis and treatment of their patients. Data are also essential when assessing the quality of care administered by the hospital staff.

Quality assessment is a hospital-wide function. It applies not only to patient care but is also incorporated in other departments, such as patient accounts, housekeeping, and security and food service. Whether to validate the accuracy of an employee's work or to assess the quantity of work performed in a designated period of time, data serves as the primary means of performance evaluation. As health care costs keep rising and as patients are faced with higher co-payments and lower deductibles, patients will demand better quality for their medical dollars. As the crisis in health care continues, health care facilities will need quality data to justify expenditures and to demonstrate quality of care. A greater emphasis will be placed on quality assessment and improvement. TQM (total quality management) and CQI (continuous quality improvement) are two processes that orginated in the manufacturing and business sectors and have been adopted by healthcare entities to maximize efficiency and quality of care. Data collected by the health care facility will become increasingly important in quality assessment and in demonstrating the need for facilities, staff, equipment, and services.

G. SUMMARY

1. Statistics is a broad term and makes use of data. Descriptive statistics and inferential statistics are representative types of statistics.
2. Data is information. Similar information gathered about a group can be organized in a data base. The processing of the information collected is referred to as data processing. Data terms include discrete and continuous data, grouped and ungrouped data, nominal and ordinal data, and computerized data. A great variety of data can be collected, including dates, test results, diagnoses, procedures, and treatments.
3. A population includes an entire group. A sample is a subset of a population.
4. A variable is something that can change. A constant assumes only one value.
5. Variables are subdivided into qualitative and quantitative variables.
6. Data which reports disease statistics is referred to as morbidity data; mortality data reports death statistics.
7. Demographic data is data on human populations and incorporates factors such as age, sex, ethnicity, income and health status of its members.

8. Vital statistics references data on human events. The primary concern of vital statistics is the individual and the major events in an individual's life—birth, death, marriage, divorce, and disease.
9. Abbreviations are used for the sake of brevity and are especially common in the health care arena. The abbreviations most commonly used in statistical computations are listed in this chapter.
10. Data has many uses and the proper collection and interpretation of data will become increasingly important as health care reimbursement dwindles and emphasis on quality assessment increases.

H. CHAPTER 1 TEST

1. Indicate whether the data represented in each of the following examples is part of a population or a sample:

a.	Twenty-five cases of TB have been reported in the past year and a patient care evaluation study is to be carried out using data from all 25 cases.	Population	Sample
b.	Sixty gastroscopies have been performed during the past two months and a study is to be carried out regarding various variables. Twenty-five of these cases will be reviewed.	Population	Sample
c.	A total of 388 chest x-rays were performed during the past month. A quality control review is to be carried out on 10% of the group.	Population	Sample

2. Indicate the terms for:

a. A value that can change _____

b. A value replaceable by only one number _____

3. For each of the following, indicate if the data is nominal or ordinal.

a.	Educational level	Nominal	Ordinal
b.	Fitness status based on a rating scale	Nominal	Ordinal
c.	Medical record number assigned by the hospital	Nominal	Ordinal
d.	License plate number	Nominal	Ordinal
e.	Placement (finish) in the 50-yd dash	Nominal	Ordinal

4. Indicate whether the following represent quantitative or qualitative variables:

a.	Type of insurance	Quantitative	Qualitative
b.	Place of birth	Quantitative	Qualitative
c.	Number of hospital admissions	Quantitative	Qualitative
d.	Number of chemotherapy treatments	Quantitative	Qualitative
e.	Blood pH	Quantitative	Qualitative
f.	Exercise engaged in for fitness	Quantitative	Qualitative
g.	Urinalysis glucose level	Quantitative	Qualitative
h.	Condition of patient at time of discharge	Quantitative	Qualitative

5. Indicate whether the data associated with the following are discrete or continuous data:

a.	Birth weight	Discrete	Continuous
b.	Cost of hospital stay	Discrete	Continuous
c.	Number of times a patient sees her physician during the year	Discrete	Continuous
d.	Number of children in a family	Discrete	Continuous
e.	Platelet count	Discrete	Continuous
f.	Deaths reported in November	Discrete	Continuous
g.	Minutes needed to walk a mile	Discrete	Continuous

6. Indicate the term for the type of data on which:
   a. Mortality statistics are computed. _____
   b. Morbidity statistics are computed. _____

7. Fifty students completed a medical terminology course at State University. The scores on the final exam were recorded as follows:

93	75	98	74	77	54	78	57	72	99	86	63	72	77	70
44	66	73	48	82	84	50	66	81	68	95	90	91	60	72
71	88	44	38	92	67	75	82	81	66	70	90	55	97	72
74	84	55	49	100										

   a. Rank the individual scores from best to worst.
   b. List each individual score only once and place a tally mark after each score.
   c. Using the grouping below, place a tally mark after each interval for each of the final scores.

98–100	82–85	66–69	50–53
94–97	78–81	62–65	46–49
90–93	74–77	58–61	42–45
86–89	70–73	54–57	38–41

8. Identify the following abbreviations:
   a. NB
   b. Σ
   c. A&D
   d. A&C
   e. DOA
   f. IP
   g. LOS
   h. ICU
   i. >