BASIC BIOSTATISTICS IN MEDICINE *And* EPIDEMIOLOGY

Alfred A. Rimm
Arthur J. Hartz
John H. Kalbfleisch
Alfred J. Anderson
Raymond G. Hoffmann

Basic Biostatistics in Medicine and Epidemiology

APPLETON-CENTURY-CROFTS/New York

80 81 82 83 84 / 10 9 8 7 6 5 4 3 2 1

Prentice-Hall International, Inc., London
Prentice-Hall of Australia, Pty. Ltd., Sydney
Prentice-Hall of India Private Limited, New Delhi
Prentice-Hall of Japan, Inc., Tokyo
Prentice-Hall of Southeast Asia (Pte.) Ltd., Singapore
Whitehall Books Ltd., Wellington, New Zealand

Library of Congress Cataloging in Publication Data
Main entry under title:

Basic biostatistics in medicine and epidemiology.

Bibliography: p.
Includes index.
1. Medical statistics. 2. Biometry.
3. Epidemiology—Statistical methods. I. Rimm, Alfred A., 1934–
RA409.B35 1979   610'.1'82   79-20011
ISBN 0-8385-0528-7

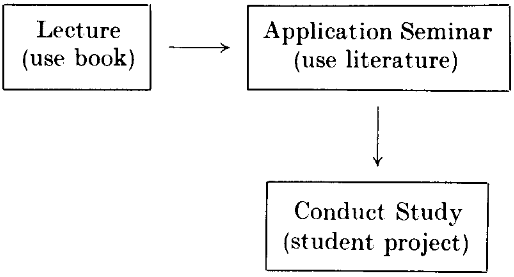
Text design: Madalyn Stone
Illustrations and page layouts by K & S Graphics, Island Park, N.Y.

PRINTED IN THE UNITED STATES OF AMERICA

Preface

This book grew out of a set of notes for an elementary course in biostatistics for medical students at the Medical College of Wisconsin. We believe the student should have a firm grip of a few critical concepts in statistics, and to emphasize them we have devoted whole chapters to a single concept. That is why the standard error of the mean (SEM), the normal distribution, and the t test are each discussed in a separate chapter. We think it is essential to teach first the difference between the standard deviation and the SEM. This gives the student the building blocks for understanding inferential reasoning and the purpose of the various statistical tests.

We have successfully used the material in this book for our course which uses the following teaching model.



The lectures supplement the notes and are interspersed with personal anecdotes to keep the students awake. Each lecture is followed by a small group seminar where a paper from the literature is used to demonstrate, with a real live example, the material covered in the previous lecture. Toward the end of the course the students are required to conduct a series of street interviews using a uniform data collection form. This is the reason for chapters on questionnaire design and protocol preparation. The data from all the interviews are pooled and groups of 15 students are required to analyze the data and prepare a scientific report. Each group receives the same data set. This research project gives the student the opportunity to utilize the methods they have learned. Just as important, this exercise demonstrates that different groups of students (investigators) working with the identical data obtain different results.

The papers from the literature demonstrating a particular procedure or method are listed under References at the end of the book. The references for each chapter are grouped according to whether they are applied or conceptual. The references listed under the general heading "conceptual" give the reader direction in pursuing an in-depth understanding and derivation of the procedures or techniques presented in each chapter.

We advise the instructor using this text to use the "application" papers listed for each chapter instead of contrived problems, which in many texts appear at the end of the chapter. Many interesting questions about these papers can be asked by the instructor. Because the questions relate to a real situation, it is easy to maintain the student's interest. The discussion of a paper does not have to be confined to the procedure or method in the particular experimental setting. Also, the author's interpretation of the data can be discussed in light of the design, sample, and analytic methods. Students find it interesting to learn that there are many different ways the author could have designed the experiment. The students usually enjoy suggesting and defending alternate study designs.

Our goal is to educate students to a philosophy of how to read the literature with a mature eye. Our philosophy is: don't criticize research work because it is not perfect and don't accept everything because it is in print. We try to teach the student to weigh the investigator's report and decide the merits of the findings on what seems reasonable in the particular experimental setting. Though this seems vague, it is what most researchers do when sizing up a paper. It is not easy to approach a subject such as biostatistics, with all its exact formulae, as if it were a course in the philosophy of extracting knowledge from scientific experiments. Yet we know that one set of data may be used to demonstrate opposing points of view. Since there is considerable, valid latitude in making underlying assumptions concerning an experiment, the two conflicting interpretations from one data set may be equally valid. Resolution of such problems usually requires accumulation of additional data and experience. The student should recognize the time and work involved in getting a single good answer to a valid scientific question. The biological sciences grind forward slowly, mainly because the secrets locked in nature are closely held, and the grinder is *man* with all of his inadequacies.

In the preparation of this book the authors acknowledge the assistance of students in the past (and future) who have pointed out errors. Also, the authors thank Ms. Barbara J. Dockery for her assistance in preparing the manuscript.

All royalties from this book are placed in a research scholarship fund for needy medical students at the Medical College of Wisconsin.

Contents

Background

I

I

Introduction

The principal purpose of this book is to create a formal framework for the physician when exercising "judgment" about scientific data. Since the physician is expected to make decisions and offer advice, he must develop and utilize his "professional judgment." Expression of judgment may be thought of as the precipitate of a process that mixes and filters ideas, experience, and knowledge. There are certain biostatistical scientific tools and concepts that constitute the fabric of the filters. This fabric is woven together by scientific logic and principles. Thus, an understanding of biostatistical tools and concepts are similar to equipping one with a filter for separating the clear signals from the background noise.

When the physician reads a medical report, he should employ scientific principles for evaluating it. For example, after reading a report where 60 patients had relief of symptoms after a course of drug X, what questions need answering before he/she prescribes the drug? This book will present the groundwork for asking the right questions, understanding the concepts for obtaining an answer to the questions, and the philosophy for evaluating the answer. The effectiveness of the physician will be greatly enhanced if decisions can be made on the basis of a scientific evaluation of the available evidence rather than on the basis of a whim, blind faith in a name, or clever advertising.

Furthermore, this book will permit you, the student or physician, to read undigested basic research articles in clinical journals. When reading these articles you must wear your scientific "hat" to understand them. Your practitioner "hat" must be worn to apply your knowledge. It is important to recognize that this text describes a scientific methodology that forms a cornerstone for clinical judgment.

The scientific methodologies discussed in this book generally fall under the heading of *statistics*. Since the methods discussed will be used in evaluation of biologic problems the more specific term *biostatistics* is used, which includes both the statistical methods and the field in which they are employed.

Statistics is just one of the many subdivisions of mathematics. Mathematics is a language with all its symbols and syntax. Statistics in a sense is a dialect of this language. The statistical formula may be likened to declarative sentences. They make a statement, and it can be proven through appropriate algebraic manipulation that the sentences (or formulas) are true. It would seem, therefore, that if mathematics hands down true statements in the form of formulas for analyzing data, then applied statistics reduces to the routine plugging of data into appropriate formulas. There is one situation where this is true—the perfect world. Unfortunately, experiments are performed and data are collected in the real world, which is far from perfect. What helps to complicate matters even more is that the marriage of statistics to biology is like many marriages where opposites attract. Whereas mathematical formulas underlying the statistical tools are exact, the science of biology is constantly changing. The solution of appropriate biologic problems with statistical tools is more of an art than a science. Science embodies an orderly set of facts and an art embodies a set of guidelines that are employed in creative ways. The discipline of biostatistics utilizes the exact mathematical formulas only as guidelines in dealing with experiments in biology and medicine. The art of medicine has many parallels with the art of biostatistics. Unfortunately, the clinician views biostatistics as an exact science, where a programmed computer can grind out all the answers that the biostatistician supplies.

A recent incident experienced by one of us (A.R.) helps to demonstrate the confused view clinicians and medical investigators have about the discipline of biostatistics. After presenting a talk to a group of cooperating clinicians where I had pooled and analyzed their data, one of the clinical investigators asked the question, "How do I go about selecting a biostatistician to help me with my work? I have gone to two different experienced biostatisticians seeking advice how to conduct a study, and I have gotten different answers!" My answer was quite simple: "I have the same problem in selecting a physician when I am sick."

The biostatistician and physician have many things in common. Each spends considerable time diagnosing a problem, each plans a strategy for dealing with the problem, and each employs tactics for solving the problem. Each evaluates the outcome to determine whether the strategy and tactics were effective. If not, then the process is repeated, until through trial and error the problem is solved or gets shelved as a problem without a solution. It is these shelved problems that both frustrate and challenge us. There is a great satisfaction in taking a tough problem off the shelf and bring to bear new tactics with an experienced hand to solve it.

Just where does biostatistics fit in the bigger picture of mathematics in biology? The field of study that includes both is called *biometry, bio* standing for *biology* and *metry* standing for *measurement.*

Biometry is an interdisciplinary field in the biologic and health sciences. It interrelates biology, mathematical statistics, and computer science in the study of quantitative concepts in the life sciences. Biometry encompasses many different aspects of "number" and biology. Almost all of these activities can be grouped into three subfields—biomathematics, biostatistics, and computer science.

BIOMATHEMATICS

Mathematical biology is concerned with the development of mathematical models for various biologic phenomena. Whereas mathematical physicists have constructed models for phenomena in the physical sciences for well over 200 years, only relatively recently have mathematical theories been applied to biology. One of the first applications was the mathematical theory for blood circulation, which corresponds to observed dilutions of injected dyes. Another early example was the theory of interaction of species for food competition. With recent findings in the biologic sciences at the micro level, mathematical theories for the transport of ions through membranes and probability models for cell division, growth of cell populations, and cell differentiation have appeared. Other mathematical models attempt to describe the behavior of the peripheral and central nervous systems, pharmacologic action of drugs, and the metabolism of substances such as glucose, iodine, and calcium. Still other models attempt to express pulmonary function and the theories of learning and vision. The successful application of mathematical models to predict the movements of the planets in the solar system is an excellent demonstration of the potential for mathematical systems.

BIOSTATISTICS

Biomedical statistics is concerned with the strategy for planning experiments, designing experiments, collecting data, classifying entities, analyzing data, and interpretating results in the life and health sciences. The tools or techniques for these activities constitute the discipline of biostatistics. Essentially, biostatistics is a methodologic science; its bread and butter is the application of methods to solve problems. A central concept is inferential reasoning. Inference is concerned with inductive reasoning, that is, generalizing to a population after observing only a sample. During the past 50 years, a large body of mathematical theory for scientific inference has been developed to provide a basis for generalizing the framework to test hypotheses. Methods are available for permitting interpretation of data from small experiments. Refined techniques permit a scientific approach for designing experiments. The application of existing procedures to problems involving patient care and cooperative efforts among clinical investigators requires an understanding of both the statistical techniques and the biologic problem being studied.

COMPUTER SCIENCE

Computer science as applied to biologic information is concerned with data processing. In addition to data analysis, it includes storage, retrieval, and transmission of information. These four aspects are combined in many on-line activities of data processing. These projects include discrete and continuous signals from biologic systems that are received and immediately analyzed and returned to the biologist or physician for the monitoring of the process. An example is the monitoring of a patient in the operating room. The patient's heart rate, systolic and diastolic blood pressures, and electrocardiogram, as well as other variables, are continuously measured, transmitted to the computer, analyzed, and reported to the operating room team. Another application is hospital medical record systems. An increasing emphasis is being given to this area of health care delivery systems so that the system and its participants may be evaluated and improved. The computer is frequently used to model or simulate physiologic systems. Changes in the conditions of the system can be simulated and evaluated by employing different strategies for solving a problem. The computer is used for data analysis, and there are many different library programs available for employing a particular statistical technique. Biostatisticians utilize the computer as a tool for data management and analysis.

ROLE OF THE BIOSTATISTICIAN

The biostatistician usually participates as a member of a research team. The medical researcher, be he surgeon, internist, or psychiatrist, presents his problem to the biostatistician. Together they determine the most meaningful experimental design to answer the question(s) being raised by the researcher. After the data are collected, the medical researcher again consults with the biostatistician for analysis of the data and for meaningful interpretation of the results.

The formal structure of the scientific approach, which is always in the "back of the mind" of the researcher and biostatistician when undertaking a project, is given in Fig. 1.

The scientist sets out to test his ideas, not to prove them. In his mind's eye, he prefers to think of the universality of his idea. However, the hurdles in the real world, where ideas are tested, limit experimental conditions, which in turn restrict generalization.

It is rarely possible in medical research to adhere 100 percent to a study design. Frequently experimental subjects drop out of a study, or data are unavailable, or records are incomplete, etc. These problems create difficulties in the evaluation of the data. Sometimes it is difficult to determine when a study did not adhere to its original design. Defects are sometimes cleverly suppressed by the authors.

When reading the clinical or basic science literature, we must remember that the process of the propagation of knowledge is contaminated, because human beings (with all their inherent defects) play a critical role in the process. If you remember everything you read, you will do well. If you believe everything you read, you haven't learned anything.

In summary, this book emphasizes the methods for medical research by focusing on:

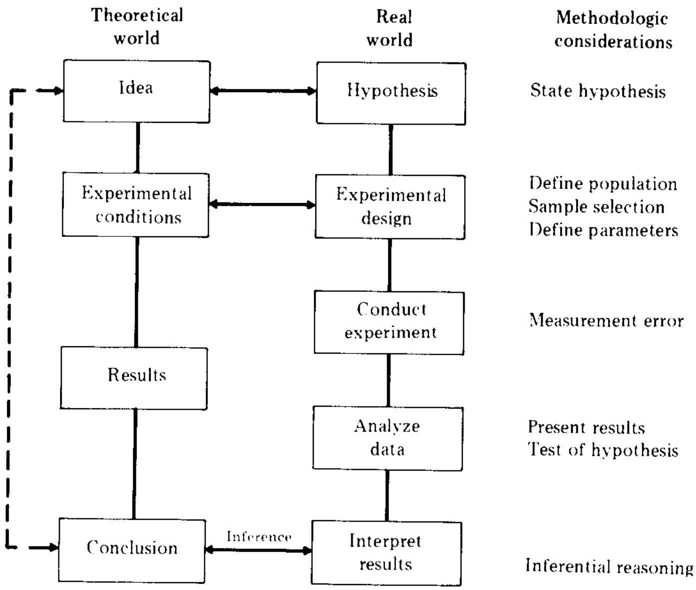


Figure 1 Conceptual diagram of the relationship between the scientist's theoretical world and his real world.

1. Sampling
2. Design of experiments
3. Analysis of data
4. Interpretation of results

The following article by Schor and Karten points out the need to know these methods for today's physician.

Statistical Evaluation of Medical Journal Manuscripts

STANLEY SCHOR AND IRVING KARTEN

Contributors of scientific communications to medical journals are responsible for the research designs of their studies, the applicability of the statistical tests used, and the validity of the conclusions drawn. This is the policy of many editorial boards. The contributors in many cases have had very little, if any, training in research methods. Some, before the investigation is begun, seek the advice of statisticians whose specialty is designing experiments so that valid conclusions may be drawn. In other instances the statistician is involved in each stage of the study. In most instances, however, research projects have not had the benefit of sound statistical advice.

The purpose of this study is to investigate the effect of this lack of statistical planning and evaluation and to present a program that may be of some help in correcting the deficiency.

MATERIAL AND METHODS

From a list of 67 most frequently read medical journals, the following ten were selected: *Annals of Internal Medicine; New England Journal of Medicine; Archives of Surgery; American Journal of Medicine; Journal of Clinical Investigation; American Journal of Diseases of Children; Surgery, Gynecology and Obstetrics; Archives of Neurology; Archives of Pathology;* and *Archives of Internal Medicine.* These periodicals are certainly not representative of all medical journals, but the list contains those that are read more frequently and considered by many physicians to have excellent reputations. Three issues of each journal were selected at random from the first three months of 1964. All communications not classified as case

From the Department of Biostatistics, American Medical Association, Chicago. Mr. Karten is a medical student at the Chicago Medical School.
Reprinted with permission.

reports by the journals were reviewed and evaluated. Additional issues were selected if the total number of studies read did not reach 25. Once an issue of the journal was selected and a contribution evaluated, the entire issue was read. This gave rise to differences in the numbers of articles evaluated in each journal.

Each communication was subjected to an abbreviated but intensive critical reading by a competent biostatistician with experience in reviewing scientific publications. The reviewers did not try to assess the relative merits of various methods of performing the experiment; they did not try to determine if the investigator used the best possible design; they did not attempt to discover whether all the important information was extracted from the data. In short, they did not compare the actual experiment with the manner in which they, themselves, would have designed it if they had been consulted in advance.

The only requirement made was that the conclusions drawn were valid in terms of the design of the experiment, the type of analysis performed, and the applicability of the statistical tests used or not used.

Since some communications published were described by the reviewers as case studies even though they were not listed as such, all reports were divided into two groups: analytical studies and descriptions of cases (case studies) (Table 1). A few reports were classified as "narrative" and, as such, did not fall into either of the two categories. These were not included.

Since the types of errors which could give rise to invalid conclusions could not be anticipated, a completely exhaustive and mutually exclusive list of errors could not be set up beforehand. Instead, the errors were listed as they occurred.

The frequency with which each of these types of errors occurred by individual journal was then obtained. The studies were grouped into three

Table 1 Distribution of Case Descriptions and Analytical Studies by Journal

Journal	Number Read	Case Descriptions		Analytical	
		No.	%	No.	%
E	35	3	8.6	32	91.4
J	30	10	33.3	20	66.7
B	26	10	38.5	16	61.5
A	28	13	46.4	15	53.6
D	27	14	51.9	13	48.1
G	33	18	54.5	15	45.5
H	31	17	54.8	14	45.2
C	29	18	62.1	11	37.9
F	26	19	73.1	7	27.0
K	30	24	80.0	6	20.0
Total	295	146	49.5	149	50.5

categories on the basis of the validity of the conclusions in terms of the design of the experiment and the type of analysis performed: those which were considered acceptable, those which should have been rejected because even major revisions could not make them acceptable, and those which should have been revised before publication on the basis of the errors found. Twelve types of errors were encountered.

1. Absence of a control group when a control group was necessary to obtain valid conclusions. This is somewhat self-explanatory, but an illustration may be helpful. If an investigator stated that 40 percent of the patients in a hospital with a particular disease were female and concluded from this that the disease was related to sex or that it was more prevalent among males than females, without any regard to the sex distribution of other patients in the hospital, the communication was placed in this category.

If an investigator wishes to know whether a particular disease is associated with sex, he should calculate the ratio of the number of females with the disease to the number of females in the population. He should then calculate the ratio of the number of males affected to the number of males in the population, and compare the two ratios. If in his hospital 5 percent of the males and 10 percent of the females were affected, then the disease is twice as prevalent among females if all other pertinent characteristics are similar. The fact that 40 percent of the patients with the disease were females is irrelevant.

Another example of a communication with this type of error is the study in which the investigator observes a characteristic in a group of people and then administers some type of treatment and observes a change. He concludes that the change was due to the treatment without using any control group to determine if the change also would have occurred without the treatment.

2. Use of measures of sensitivity without measures of specificity in evaluating a treatment, drug, or diagnostic procedure. "Sensitivity" is defined as the percentage of cases that are detected by a given procedure. "Specificity" is defined as the percentage of "noncases" identified as noncases by the procedure. In some reports evaluating diagnostic tools of one type or another, only one of these is measured. An investigator might claim, for example, that his new diagnostic technique identifies 98 percent of all the cases of a given disease in his sample. From this he concludes that it is a good diagnostic tool. It is imperative, however, that he also observe the number of times the diagnostic tool identifies a person without the disease as having the disease (false-positives). Certainly, if this is not done one can develop an excellent test for heart disease by counting a patient's ears. Everyone who has two ears has heart disease. This would miss practically no heart disease case. It is still not a good diagnostic tool, however, because it has very low specificity. It identifies noncases as cases too often.

3. Improper use of statistical techniques. Any study in which the investigator applied statistical tests which were not justified by the data or the data-collection procedure was included in this category. Examples