Terezis, L. N., W. C. Davis, and F. C. Jackson. 1963. Carcinoma of the colon associated with inguinal hernia. *New Engl. J. Med.*, **268**:774.

Truett, J., J. Cornfield, and W. Kannel. 1967. A multivariate analysis of the risk of coronary heart disease in Framingham. *J. Chron. Dis.*, **20**:511–524.

Vessey, M. P., and R. Doll. 1968. Investigation of relation between use of oral contraceptives and thromboembolic disease. *Brit. Med. J.*, **2**:199–205.

Wijsman, R. A. 1958. Contribution to the study of the question of associations between two diseases. *Human Biology*, **30**:219–236.

Yerushalmy, J., Statistical considerations and evaluation of epidemiological evidence, in *Tobacco and Health* edited by G. James and T. Rosenthal. (Springfield, Ill.: Charles C Thomas, 1962), pp. 208–230.

Chapter 12

How to Carry Out a Study

Many health-care professionals wish to conduct a modest clinical or epidemiologic study. Hoping to answer one or more interesting questions, they find themselves in a good position to collect and analyze some appropriate data. However, to someone without previous research experience, the task often appears awesome, and it is not at all clear how to proceed.

This chapter is written as a general guide for the novice who wishes to carry out such a study. Obviously, each research project and each study setting presents unique problems which cannot be dealt with here. What will be presented is a general approach which emphasizes the practical difficulties that are frequently troublesome to the beginner.

Defining the Problem

The first step—and one of the most difficult ones—is defining the problem and choosing the question or questions to be answered.

There is a tendency for the novice at research to ask questions that are diffuse or vague. Instead, the problem must be stated in terms of clear, simple, answerable questions.

An example of a vague unachievable goal for a specific study would be *to elucidate the role of psychological factors in coronary heart disease.* It is not clear whether the "role . . . in coronary heart disease" refers to causation of the disease, outcome of disease, the patient's attitude toward the disease, or something entirely different. Furthermore, both "coronary heart disease" and "psychological factors" are very broad terms. Better, because they are clear and answerable, are specific aims or questions such as, *Determine the proportion of patients with myocardial infarction who develop severe emotional depression during hospitalization. Do attacks of angina pectoris occur more frequently during periods when patients are anxious?* Or, *Is there an increased risk of sudden cardiac death within a year after the death of a spouse?*

Intimately involved in the asking of vague, overly broad questions is the tendency to be too ambitious. The new researcher wishes to make important discoveries and solve big problems. These unrealistic expectations can only lead to failure and disappointment. For the most part, medical science progresses gradually by very small steps. So much of health care is based on tenuous evidence and incomplete knowledge that a careful study of a simple question will be a worthwhile contribution, of which any scientist should be proud.

Reviewing the Relevant Literature

Once a problem has been selected, the scientific papers describing previous related work should be read carefully. In addition to learning what is already known about the question, the investigator will become familiar with problems that others have faced, using various study methods. One should be especially alert for related variables which can be measured or controlled in the planned study so that embarrassing spurious correlations can be recognized or avoided. For example, no investigation of a possible etiologic factor in lung cancer would be respectable if smoking habits were not measured or taken into account.

The usual result of a literature review will be a realization of how *little* is known about the particular topic one wishes to investigate. Seemingly authoritative statements and accepted medical doctrines, perpetuated through textbooks and lectures, often turn out to be supported by the most meager of evidence, if any can be found at all! For example, my own experience in reviewing the literature for an epidemiologic study of gallbladder disease was an inability to find any evidence for the "fair" and "forty" parts of the doctrine that persons who are "fair, fat, and forty" are especially prone to gallstones. Indeed, the study did not confirm these traits as predisposing factors.

Many other examples could be mentioned of beliefs that are based on little or no evidence or on the results of poorly conducted studies. In these instances, a literature review will provide encouragement for proceeding with the proposed study. On the other hand, if it is evident that the question has already been well answered, a related problem may come to mind—one that can be studied just as well.

Preparing a Protocol

The next essential step is the preparation of a study protocol. Even though the beginning investigator may feel that he has clearly in mind what he plans to do, it is extremely important to set down the plan in writing.

A written protocol serves three major purposes. First of all, when one writes the protocol, ideas and procedures must be clearly defined and spelled out. Usually the plan in one's mind is not as clear and logical as was hoped, and the gaps and flaws are easier to recognize and correct when the plan is seen on paper. Secondly, a written protocol can be studied by anyone whose advice is desired or whose approval is required. Thirdly, any person working to carry out the study, even the investigator himself, may forget some method or procedure to be followed. The written protocol constitutes a permanent record that can be referred to, so that methods do not change unnecessarily during the conduct of the study.

Some persons have such an abhorrence of writing that the preparation of a protocol is an almost insurmountable obstacle to

carrying out a study. If so, it is probably better just to quit at this point, since even if some data are collected and analyzed, the results will probably never be written up and no one else can adequately study the findings. Another alternative for a nonwriter with a good idea for a research project is to team up with a co-investigator who is willing and able to write the protocol and the final report.

Contents of the protocol Research grant applications may require strict adherence to prescribed contents arranged in a particular order. For example, a recent communication from the U.S. National Institutes of Health listed the following required elements for a grant application.

Broad statement of objectives
Detailed budget for the first year
Budget estimates for subsequent years
Biographical sketches for all professional personnel
Research Plan

- **A** Introduction
  - **1** Overall objective or long-term goal
  - **2** Background: significant previous work and current status of research
  - **3** Rationale behind the proposed approach to the problem
- **B** Specific aims
- **C** Methods of procedure—methods used, data to be collected, how the data will be analyzed and interpreted, possible pitfalls and limitations, tentative schedule
- **D** Significance of the proposed work
- **E** Facilities available
- **F** Collaborative arrangements, if any
- **G** Appendix—various detailed descriptions, letters confirming proposed collaboration, etc.

A protocol prepared for local use may be shorter and simpler but should contain at least the following elements, unless there is good reason for omitting any.

1. A brief statement of the specific question(s) to be answered and/or the specific aim(s) of the study

2. Background and significance of the study. This should be a pertinent nonrambling discussion of what is known and not known about the problem and why the proposed study is worthwhile or important.
3. Methods. Included should be a description of the study subjects—how they are to be selected and how many there are likely to be. The data to be collected and the methods for collecting them should be described. Uniform criteria for diagnosis of disease and for decisions as to the presence or absence of a characteristic or outcome should be listed. Data analysis methods should also be presented, preferably with some sample blank tables showing how the data will be organized. Plans for safeguarding the rights and welfare of the subjects and the method of obtaining their informed consent (if needed) should be explained.
4. An approximate time schedule for carrying out the various aspects of the study.
5. A budget, if financial support is being requested, with explanation of any personnel and other costs whose requirement is not obvious.

Consultation

After a draft of the protocol has been written, it is wise to seek some expert consultation before proceeding any further. Many potential problems and difficulties will be quickly spotted by knowledgeable persons reviewing the protocol and discussing the proposed research.

It should be no reflection on one's intelligence and skill to ask for advice. No one can foresee all the problems that may develop in his own study. A consultant will respect the investigator who draws up a protocol as well as he can and then admits that he is fallible.

Help can come from persons in a number of disciplines. An experienced investigator who has worked in the area to be studied can perhaps provide the most comprehensive view of the problem. A clinician who specializes in the area of study will often provide some fresh insights into the subject matter derived from experience with patients and from familiarity with the current literature. Epidemiologists and, particularly, biostatisticians are professionally concerned with study design and data analysis and can provide guidance on

these aspects of the study. The choice of appropriate statistical tests and the determination of whether or not the proposed sample size is adequate to obtain meaningful information, are of particular concern to the biostatistician.

The protocol should now be revised taking into account the suggestions of the consultants.

Presenting the Study Plan to Other Key Individuals

At this time the investigator should inform all the responsible persons whose approval or cooperation is either required or desirable. Proposed research in medical or academic institutions should be presented to appropriate departmental heads and/or hospital administrators. Often there will be a committee specially designated to review and approve of studies. Epidemiologic studies in the community should be described to local health officials and to the medical society.

In addition to gaining the required approvals, the investigator may receive valuable practical suggestions and other assistance from these individuals, such as introductions to physicians who may permit the study of their own patients. The investigator may also learn of other similar or related research that is under way. Cooperation with other investigators may help avoid duplication of effort and may lead to sharing of resources and, possibly, even of data.

Data Collection Methods

The data to be collected—whether by observation or interview of subjects, by chart review, laboratory tests, or however—must be recorded in a systematic and orderly manner. The usual method of bringing order into the data-recording process is by the use of standard forms. Careful attention to preparation of a form, even if only a few items need to be recorded on it, will save the investigator from much trouble and grief later on.

One or more forms will be used for each study subject. Each form should provide space for identification of the subject and for recording the necessary data about him.

If mechanical or electronic data processors are to be used for analysis, the format for recording data on the form should meet the

requirements of these devices. Each unit of information must be recorded in a particular space on each form. Each space is ordinarily assigned a column number to correspond with the column on a punch card to which that unit of information will be transferred. Currently, most data processing equipment accepts information from 80-column punch cards, but because of local variations, the investigator should seek advice from data processing personnel at his institution before drawing up the form.

For recording quantitative information, specific spaces or boxes should be designated so that the same digit (e.g., the "ones," "tens," or "hundreds" digit) is entered into the same space on each form, and the location of the decimal point is uniform. If the value to be recorded is relatively small and does not require all the assigned spaces, zeros should be written in the spaces to the left, which would not otherwise be filled in. Adequate spaces should be provided for all possible values of any particular measurement and for recording that the value is unknown. Special instructions for recording each measurement may be located on the form itself or in an accompanying manual.

For example, suppose an investigator wishes to record the serum-glucose level at admission to the hospital and he provides three spaces on his form, to be transferred, say, to columns 20–22 of the punch card, as follows:

Serum glucose (mg/100 ml) ☐☐☐

Cols. 20–22

At first glance this may seem adequate, but consider what might go wrong if a research assistant tries to use these spaces for three patients, one with a value of 72, one with 1,021, and one for whom the test was not done. Without special instructions to use the two boxes on the right for two-digit numbers, the value of 72 might be recorded as [7][2][ ], which will be treated by the computer or card sorter as 720. In recording the value 1,021 the naïve research assistant might well write 1 [0][2][1] not realizing that what is outside the three boxes will be lost in data processing. The investigator should have anticipated the possibility of the occasional extremely high value for a patient suffering from diabetic acidosis and provided

four boxes instead of three. If there is good reason to limit the spaces to three, another less-satisfactory alternative is to make a rule for high values such as "Code 999 for values of 999 or greater. Write actual value below." The value can then be referred to if needed. However, the computer will not be able to compute an accurate mean if 999 is always substituted for greater values.

For the patient with no glucose determination the research assistant may leave the space blank. But 6 months later when the data are to be analyzed, and the research assistant has moved to another city, the investigator will not be sure whether the blank spaces represent an unknown value or whether the assistant forgot to fill in the spaces. It is better to indicate "test not done" with a particular number that could not represent a possible value of the variable. Consideration of these potential problems leads to the improved version of the portion of the form for recording serum glucose as follows:

Serum glucose (mg/100 ml) ☐☐☐☐
Cols. 20–23

(Record one-, two-, and three-digit numbers
as far to the right as possible, and
fill in the left boxes with zeros.
If test not done, record 9999)

Qualitative data, such as diagnostic categories, or "yes" or "no" responses, usually require the assignment of code numbers to each response if data processing devices are to be used. Consider marital status, for example. Without coding, a data collection form might show marital status as follows:

Marital status (check appropriate category)

☐ Single ☐ Widowed
☐ Married ☐ Divorced
☐ Separated ☐ Unknown

The responses could be coded into a single digit if a number

were assigned to each category. The digit could be recorded in a space or box on the same sheet or onto a separate code sheet. For example, note how marital status can be coded into one digit to be transferred to, say, Col. 17 on a punch card.

Marital status (enter appropriate number into box) ☐ Col. 17

1 Single **4** Widowed
2 Married **5** Divorced
3 Separated **6** Unknown

Precoded forms permit the correct category to be marked and coded automatically. For example:

Marital status (circle number next to appropriate category)

Single	1	Col. 17
Married	2	
Separated	3	
Widowed	4	
Divorced	5	
Unknown	6	

There are advantages and disadvantages to each type of form. Some general principles to consider are:

The less rewriting or transcribing of data that is needed, the less chance for error.

The less complex the form, the less chance for error.

Most physicians and other professionals neither like to code nor do a good job of coding. If such individuals are to record data, it is often necessary to design a form they will use, and pay someone else to do the coding.

In preparing to record qualitative data, a category should be provided for every possibility except the very rare ones. Writing of additional information on the form in longhand should be kept at a minimum because this sort of information is difficult to analyze and relate to the other variables. Consider, for example, a study of factors related to adverse reactions to anticoagulant drugs. One item of

information that will be desired about each patient is the medical condition for which the anticoagulant is given. The investigator could set up his code sheet as follows:

Condition for which anticoagulant was given:_____

_____

_____

However, he might later find it difficult to summarize these data and combine patients into categories. Using his clinical experience to anticipate the possibilities, he would find the data easier to analyze and present by providing several mutually exclusive categories, as follows:

Condition for which anticoagulant was given: ☐☐ Cols. 32–33

01 Pulmonary embolism
02 Thrombophlebitis
03 Pulmonary embolism and thrombophlebitis
04 Myocardial infarction
05 Myocardial infarction with mural thrombosis and peripheral embolism
06 Rheumatic heart disease
07 Rheumatic heart disease with peripheral embolism
08 Atrial fibrillation or flutter
09 Atrial fibrillation or flutter with peripheral embolism
10 Atrial fibrillation or flutter with therapeutic conversion
11 Prosthetic heart valve
12 Transient cerebral ischemic attacks
13 Other cerebrovascular disease, specify _____
14 Other disease, specify _____
15 Combinations of above, specify _____

_____

Note the last three categories which involve some specification in longhand. These permit the recording of unanticipated conditions. But provision of the other common categories will reduce the need for longhand recording to a very small fraction of the cases.

When specifying categories for data-collection forms it is wise to avoid making these categories too broad. Overly broad categories lead to the loss of valuable information. For example, categories, 08, 09, and 10 above, might have been combined under a more inclusive category "atrial fibrillation or flutter," but then, important clinical distinctions among these cases could not be made without referring to the chart again. Frequently the investigator assumes that broad categories will be adequate for the needs of the study. Later on when the data are analyzed, unanticipated questions arise which could have been answered if narrower categories had been used.

Broad categories may prove especially troublesome when quantitative variables are recorded. In providing for the coding of serum glucose it might initially seem reasonable to have only 7 categories:

1 Less than 50 mg/100 ml
2 50–99
3 100–199
4 200–499
5 500–999
6 1,000+
7 Test not done

With luck, this might be perfectly adequate. However, if another investigator's study shows an important difference in findings between persons whose glucose level is less than 350 mg/100 ml and those whose glucose is 350 mg/100 ml or greater, the broad categories chosen will not permit data analysis to determine whether the breakpoint at 350 mg/100 ml can be confirmed. Furthermore, it is not possible to compute accurate means and standard deviations with the crude breakdown as shown above.

Thus it is best to record quantitative values exactly as they come from the measuring device. This allows for maximum flexibility and permits the investigator subsequently to use any grouping he desires.

Pretesting of Data Collection

No matter how carefully the data collection is planned, problems will come to light after starting. That is why it is important to pretest procedures and forms before the study formally begins.

Suppose, for example, that data for a study of cardiovascular

disease are to be collected in a mobile facility in which volunteer subjects are scheduled to pass from station to station every 5 minutes for a series of procedures. It may turn out that the electrocardiogram takes 8 minutes, on the average, instead of the planned 5 minutes. As a result subjects may pile up at earlier stations if there was no provision for a waiting area in case of delays. It may therefore be necessary to slow down the examination schedule, or provide two electrocardiographic stations, or set up a waiting area. This problem should be uncovered and solved during pretesting. If not, and the subjects have to wait or get the impression that the study is disorganized, cooperation may be seriously impaired.

Similarly, a series of interview questions may seem perfectly clear and appropriate when they are written down. Yet when study subjects are actually asked these questions they may not understand, or be offended, or give responses that were not anticipated. In a study of radiation exposure, for example, the investigator may consider it perfectly reasonable to ask, "Have you ever received x-ray or isotope therapy?" It will undoubtedly turn out that some subjects answer "yes" because they misinterpret the question to mean x-ray examinations. The question will have to be reworded and supplemented with additional clarifying questions in case of a "yes" response. Problems such as these quickly become apparent when an interview is tried out on friends and associates first, and then on some persons similar to the potential study subjects, but not officially part of the study.

Even abstracting data from charts requires pretesting. It seems perfectly reasonable to ask a research assistant reviewing hospital charts to record the patients' blood pressure at the time of admission. When the assistant looks at the first few charts it will be noted that some, but not all, patients are admitted to a ward from the emergency room, where the blood pressure was recorded by the intern. There is also a blood pressure recorded as the first of a series of blood pressures on the nurse's vital signs chart. In addition, the intern and resident on the ward each performed an initial physical examination in which the blood pressure was recorded. It is apparent that some rule will be required for selecting the blood pressure to be used, if any consistency is to be achieved. Review of a few charts will also reveal that one of the interns has recorded two

diastolic pressures, one at the muffling and one at the disappearance of Korotkov's sounds. Thus another decision is required—which one to use.

To mention other examples, the investigator may ask a chart reviewer to indicate whether the patient has a history of hypertension—yes or no. The chart reviewer will find, for a particular patient, that one physician records such a history and another does not. Which physician's history should be used? Or, on a previous hospitalization one blood pressure of 150/105 was recorded. Does this constitute a history of hypertension? Again, decisions and further clarification are needed. Or, the form was constructed so as to provide spaces for three digits for recording systolic pressure and two for diastolic, because it was forgotten that the diastolic is frequently greater than 99 mm Hg. Pretesting will reveal the need to change the form.

Data Collection

If the investigator is relying on others to collect and record the data, he should supervise this aspect of the study closely, especially during the early stages. The work of persons collecting the data should be observed, and completed data collection forms should be checked carefully. In this way, the investigator can ensure that his study plan is being followed.

Not all problems will have been discovered during pretesting. Further changes in procedures and forms may have to be made after the study officially begins. These modifications should be kept to a minimum in order to avoid inconsistencies in the data. Any changes or new rules to be followed should be recorded as additions to the protocol.

Data Analysis

Data analysis for most epidemiologic or clinical-outcome studies mainly involves sorting into categories and counting, then computing proportions, rates, means, and other group characteristics.

In order to proceed in an orderly fashion and end up with the answers that were desired in the first place, it is often helpful to draw

up some blank tables showing the format for displaying the results of data analysis as they would be presented in a final report. These tables are then filled in with the appropriate counts, rates, and so on when these results become available.

For the novice, preparing blank tables is often quite difficult, requiring a good deal of patience and self-discipline. However, the results are well worth the effort and, with experience, subsequent table-making becomes much easier.

Data analysis tables should show the results broken down by age, sex, and other pertinent variables. In addition to showing the key results that one is after, they should show the numbers upon which these results are based. For example, Table 12-1, below, showing just incidence rates, is inadequate. The counts upon which these rates are based, should also be listed, as in Table 12-2.

Table 12-1 Incomplete Table Showing Only Incidence Rates by Age and Sex (Fictitious Data)

Age	Annual incidence rate/1,000
Men	
20–29	16.5
30–39	22.8
40–49	23.4
50–59	42.4
60–69	77.1
Total	33.3
Women	
20–29	5.5
30–39	8.6
40–49	10.5
50–59	20.9
60–69	40.6
Total	16.2

Table 12-2 Complete Table Showing Incidence Rates by Age and Sex and the Numbers upon Which They Are Based (Fictitious Data)

Age	Population at risk	Number of new cases during the year	Annual incidence rate/1,000
Men			
20–29	1,572	26	16.5
30–39	1,494	34	22.8
40–49	2,012	47	23.4
50–59	1,629	69	42.4
60–69	1,077	83	77.1
Total	7,784	259	33.3
Women			
20–29	1,827	10	5.5
30–39	2,203	19	8.6
40–49	2,570	27	10.5
50–59	1,912	40	20.9
60–69	1,698	69	40.6
Total	10,210	165	16.2

Similarly, when means and standard deviations are shown, the number of persons entering into each of these computations should be given.

The actual work of data processing can be carried out in a variety of ways. The proper choice of method depends on how many subjects are involved and the complexity of the analysis. If only a few counts and proportions are to be determined for a few dozen subjects, manual counting of items on the data collection forms will be quite adequate. With increasing numbers and tabulations, the investigator may wish to sort and count his data using special cards, available at stationery stores, with holes punched near the edges. Each subject's data are recorded on one of these cards. The presence of a particular characteristic can be shown on the card by punching away the thin strip of cardboard that separates the hole from the edge. When the cards are lined up behind each other, a

long needle is passed through the corresponding hole in all cards and then lifted. The cards representing persons with the particular characteristic will drop away from the rest.

Even more helpful, and useful for studies with hundreds or even thousands of subjects, is the electric card-sorting machine. Data are keypunched onto special cards (most often with 80 columns, one for each digit). These cards are rapidly sorted by the machine into subgroups, and the number of cards in each subgroup is counted at the same time. Card sorters of this type are widely available, and the investigator can easily learn to operate one in a few minutes.

After any of the above sorting and counting methods are used, the arithmetic necessary for computing rates, means, statistical-significance tests, and so on, may be conveniently done on an electric desk calculator.

If a great number of counts and sorts have to be done on large numbers of subjects and/or if complex mathematical calculations are required, then the electronic digital computer is the ideal data-analysis tool. Voluminous data may be transferred from punch cards to a reel of magnetic tape, which may be used repeatedly for as many analyses as are desired. To use a computer, the investigator must usually obtain the services of a programmer or else learn to program it himself. If a programmer is employed, the investigator will have to explain in meticulous detail exactly what he wants. Various systems are being developed to make it easier for an investigator with little knowledge of programming to communicate directly with the computer.

Preparing the Final Report

The difficult job of preparing the scientific paper describing the study becomes a much less imposing task if the investigator writes portions of it during the course of the investigation.

The introductory section of the paper briefly outlines the problem and the purpose of the study. Note that this material has already been set down in preparing the written protocol. All that has to be done is to make any modifications that seem necessary for a final paper, or that come to mind now that data collection and analysis are completed.

Similarly, the Methods section of the paper can be readily

adapted from the protocol. It should describe exactly what was done and, in addition, inform the reader how subjects were selected for study, how many were included or excluded and the reasons for exclusion. Criteria for classifying subjects and for decisions as to outcomes should be spelled out.

Writing the Results and Discussion sections is simplified if the investigator takes advantage of the fact that the results usually appear in stages. Tables of data are usually completed one at a time. As the investigator prepares or receives each table, he should immediately write a paragraph or two describing it for the Results section and a paragraph or two discussing the implication of this result for the Discussion section. Then, by the time the data are completely analyzed, most of the writing will be done. The Results and Discussion will still have to be organized and edited, but the task of writing it will not have to be faced all at once.

In the Discussion section the implications of the study and its relation to previous work should be described. In addition, the difficulties, problems, and potential errors and biases of the present investigation should be reviewed. All investigations have some obvious limitations and others that are not so obvious. It is best if the investigator recognizes and points these out himself, before someone else does.

Importance of Good Communication

Science is a *social* process. Each investigation is related to previous work, either attempting to confirm it or, usually, to build upon it. Investigators need to know and understand what others have done and are doing. It is, therefore, an important responsibility to present a study as fully and clearly as possible.

A paper, or an oral presentation at a meeting, should be clear and simple. Jargon and unnecessarily complex or obscure terminology should be avoided. Although tables of data in a written paper should be complete, they need not be repetitious. Thus, after the basic numbers have been shown once (as in Table 12-2), they do not have to be repeated over and over again. Some relationships may be communicated most clearly by means of graphs and figures; these visual aids should be used freely.

In preparing slides or charts to accompany an oral presentation,

the temptation to crowd a lot of information into one slide should be resisted. Each slide should have only a few lines or numbers, displayed with large characters, easily seen by those in the back of the room.

Importance of Investigator Worry

Many things can go wrong and many errors can occur during a study. Therefore it is essential that, preferably, the investigator himself, or else a conscientious person responsible to the investigator, *worry about details.* The careful investigator might well adopt a questioning or even a suspicious attitude toward his study.

In addition to observing the process of data collection, as recommended above, the investigator should see to it that every data-recording form is checked carefully by someone other than the person who filled it out, to detect and correct omissions and obvious errors. Copies should be made of all completed data collection forms so that the original information will still be available if any forms are lost. Complete lists and counts of all study subjects should be maintained to provide a check against lost forms. It is surprising how often forms become misplaced or piles of punch cards fall behind a desk. Keypunching of data should be verified, which involves repeating the keypunching on a machine that detects discrepancies.

All mathematical calculations should be done twice by two different persons. The investigator should be sure to have his own work double-checked by a conscientious individual. Computer programs should be tested on small samples of data and the results compared with hand calculations.

Data tables should be checked to make sure that all the numbers are correct and add up to the totals shown. Surprising or inconsistent results should provoke redoubled efforts to check whether something has gone wrong.

Finally, it would be sad, indeed, if after all this work the resulting paper were to contain misleading typographical or printing errors. The manuscript and galley proofs should be proofread carefully.

Chapter 13

Epidemiology and Patient Care

Epidemiology is quite important in patient care. Clinical decisions are greatly affected by knowledge of the patterns of disease occurrence in populations. Some of the ways that diagnosis and treatment are, or should be, related to epidemiologic knowledge and principles will be discussed in this chapter.

Epidemiology and Diagnosis

In making a diagnosis, the physician must select from the hundreds of known diseases that one which most probably fits the patient's clinical picture. In assessing the probability of a given condition being present, the physician is strongly influenced by an awareness of what diseases are prevalent in his community at the time. During an influenza epidemic, for example, a patient exhibiting fever, headache, weakness, and myalgia would be promptly diagnosed as having influenza; whereas, with no such epidemic taking place,