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,

laboratory tests would probably be ordered to rule out other explanations for the illness. Similarly, in the United States in the 1970's, if a patient presents with congestive heart failure, diphtheritic or Chagas' myocarditis need rarely be considered.

Descriptive epidemiologic findings indicating subgroups of the population in which a disease has a low or high prevalence are also useful for diagnosis. Knowing that a patient is of a particular age or sex or occupation, or that he comes from a certain part of the country, is very helpful in narrowing down the probable diseases he or she might have. For example, if a patient has lived in the San Joaquin Valley of California, coccidioidomycosis should be strongly suspected as the disease responsible for a nonspecific lung lesion seen on his chest x-ray.

The use of epidemiologic knowledge in the diagnosis of heart disease was well described by Dry (1943) who quoted "a cardiologist of long experience" as follows:

> When I am called to see a patient with heart disease that is not of almost self-evident nature I find out certain things before I enter the room. I know whether the patient is a baby, a child, an adolescent, or an adult. If he is an adult, I find out in what age range he falls. There are certain heart diseases found commonly in certain ages and rarely found in persons of other ages. I have made my first step in probable diagnosis right then.
>
> Then, particularly if the patient is an adult, I must know whether he is male or female, for there is a sex predilection for certain diseases of the heart. That's my second step and I have narrowed the probable diagnosis down further.
>
> Next, I find out from the history what he has been exposed to. What diseases has he had? What kind of life has he lived? Has he suffered important hardships, been a rounder? Is he or she a successful, hard driving person? How much does he eat, smoke and drink? In what condition is his general health? Such questions as these narrow the problem down further. I am pretty well along in logical diagnosis by exclusion before I cross the threshold.
>
> Then I do cross it but I'm in no hurry. I shake hands with the patient, feel his pulse. I get certain impressions that way. I look at him, talk to him and size him up as a man and a doctor rather than as a cardiologist. I ask him questions about his specific complaints and continue to ply him with questions until I have the picture in my own mind of just what he has been experiencing subjectively. It is not enough to know, for instance, that he has pain in his chest or shortness of breath because either may indicate serious heart disease or a condition that is relatively innocent. Thus I have secured further background and some of what already was in my mind when I was standing in the hall has been either reinforced or refuted.
>
> Then I put my hands and my stethoscope on his chest in the course of a complete and thorough examination. Next I review the x-ray and the electrocardiogram. I ought to get every bit of evidence I can, but I honestly doubt if any of it is usually as important as the thinking I did in the hall and at the bedside before I touched the patient or had any apparatus applied to him.

Analytic Studies to Improve Methods of Diagnosis

Population studies, quite analogous to analytic epidemiologic studies, have been used to refine diagnostic methods. Just as epidemiology traditionally studies the associations between a disease and etiologic or predictive factors, the same approach may be used to study the associations between a disease and symptoms, signs, or laboratory tests. These, after all, constitute the information that is used to make a diagnosis.

Because of the current popularity of laboratory tests, one need only glance through a volume of issues of any leading medical journal to find examples of studies showing how a particular test may be used to help distinguish between persons with and without a particular disease, or between different categories of patients with the same disease. Ordinarily, the test will be performed on different patient groups, plus some "normal controls." The distributions of the test results in each of these groups are then compared. When any two distributions appear quite different and show little overlap, the test is valuable in discriminating between the two groups; that is, the test is helpful in determining whether a patient belongs to one group or the other.

The value of a symptom in distinguishing between persons with and without a disease may be investigated similarly. Two recent

population studies are examples of great interest because they showed that certain traditional clinical teachings about the relationship between a disease and a symptom are probably incorrect.

One of these studies (Price, 1963) looked closely at the relationship between various types of indigestion or dyspepsia and gallbladder disease. It had long been taught by some authorities that chronic epigastric pain, flatulence, heartburn, and intolerance to fatty and other types of foods could often be due to a diseased gallbladder. A total of 204 women, ages 50–70, were identified in one urban general-practice patient roster in the United Kingdom. Of these women, 142, or 70 percent, agreed to be interviewed concerning these symptoms. Each patient later had an x-ray of the gallbladder, by means of which 24 were shown to have gallstones or a poorly functioning gallbladder.

The relative frequency of fatty-food intolerance and of each of the other "typical" symptoms was quite similar in the groups with normal and abnormal gallbladders. Altogether, dyspepsia was quite a common symptom, afflicting about half of each group. The type of indigestion experienced by the abnormal group did not differ appreciably from that reported by normals. The author concluded that among women, ages 50–70, the presence of both gallbladder disease and dyspepsia is coincidental, and that these symptoms can not assist in the diagnosis of gallbladder disease and should not influence its treatment.

In another study of this type, Weiss (1972) analyzed data from the 1960–1962 U.S. National Health Survey to explore the relationship between hypertension and certain symptoms, long regarded as being due to this condition. Responses to questions about these symptoms on a self-administered questionnaire were studied in relation to blood pressure subsequently measured by a physician. Headache, epistaxis, and tinnitus showed no relationship to either systolic or diastolic pressure. A history of dizziness was more prevalent only in those hypertensives with a very high diastolic pressure. Fainting was inversely related to blood pressure, being reported more frequently by those with lower pressures.

It is not surprising that many physicians have accepted the teaching that gallbladder disease produces fatty-food intolerance and hypertension produces headache. First of all, these relationships, particularly the former, can be "explained" physiologically. Secondly, these symptoms are quite common; thus it is not surprising to find patients complaining of them. Nevertheless, by examining these symptom-disease relationships in general population groups, the epidemiologic approach can put them in proper perspective.

"Normal" Values

Returning again to laboratory tests and other quantitative measurements, practicing physicians and laboratory directors are in the habit of dividing the distributions of these findings into two parts, the "normal" and the "abnormal." Having a clear dividing line or "normal limit" between the two alternatives makes it easier to make decisions. If the patient is normal, he can be reassured; if he or she is abnormal, some action must be taken. Thus, it is important to understand how normal limits are arrived at.

Unfortunately, much confusion surrounds this area because the term "normal" has more than one meaning. As used above it means "good" or "desirable" or "healthy." Another important meaning is "usual" or "frequent." In this sense, it is normal for an older person to have gray hair. This says that the occurrence is common but implies nothing either way about desirability. As if these two definitions did not cause sufficient confusion, there is a third meaning having to do with the shape of a distribution curve that is often observed in studies of human characteristics. This symmetrical, bell-shaped curve is referred to as the "Gaussian" or "normal" distribution curve.

One method that has been used to define the "normal-healthy" has been to determine the "normal-usual." That is, the particular test is applied to a large population. A cutoff point is applied to one or both ends of the distribution curve so that an arbitrary small percentage, say 5 percent or 1 percent of the population, will be called abnormal. Clearly, by this method, the normal range is merely the usual range; but it is easy to drift into the view that normal-usual means normal-healthy.

This method for determining normal-healthy limits can be improved upon by finding the normal-usual values in a population that is known to be healthy. Unfortunately, the healthy group studied

is often small and select—for example, a group of medical or nursing students. Thus it is hard to be sure whether test values associated with health in these groups would also be associated with health in persons of different ages and circumstances.

Even better than studying a healthy group alone is to determine the test values in two groups, one that is healthy and one which has the disease being tested for. The result will usually be two overlapping distributions as shown in Fig. 13-1. Outside the area where the distributions overlap, a test result clearly identifies the presence or absence of disease. If a patient's value falls into the area of overlap, he has a chance of belonging to either the normal or abnormal group. Choosing one cutoff point will thus result in errors in classification; that is, there will be some truly normal individuals on the abnormal side of the cutoff point who will, therefore, be called abnormal, and there will be some truly abnormal individuals who will be considered normal.

These two types of classification errors can be expressed quantitatively in terms of the *sensitivity* and the *specificity* of a test. Sensitivity is the proportion of truly diseased who are called diseased by the test. Specificity is the proportion of truly nondiseased persons who are so identified by the test. In the example shown in Fig. 13-1 it is apparent that these two measures are inversely related to one another. Shifting the cutoff point to the left will increase sensitivity at the cost of specificity. That is, a higher percentage of sick persons will be called sick but a smaller percentage of the well will be called well. Moving the cutoff point to the right will increase specificity while decreasing sensitivity. More of the well will be called well but less sickness will be detected.

In setting the normal cutoff point, then, attention must be paid to the purpose of the test. If it is very important not to miss a particular disease which is both treatable and serious, one usually favors sensitivity over specificity, hoping to correctly identify as many cases as possible. On the other hand, if detecting a disease results in little benefit while falsely labeling normal persons as sick results in much worry and cost, specificity is to be preferred.

Unfortunately, the physician's desire for a nice cutoff point has been dealt a rather serious blow by recent epidemiologic studies, particularly in cardiovascular disease. For important coronary risk

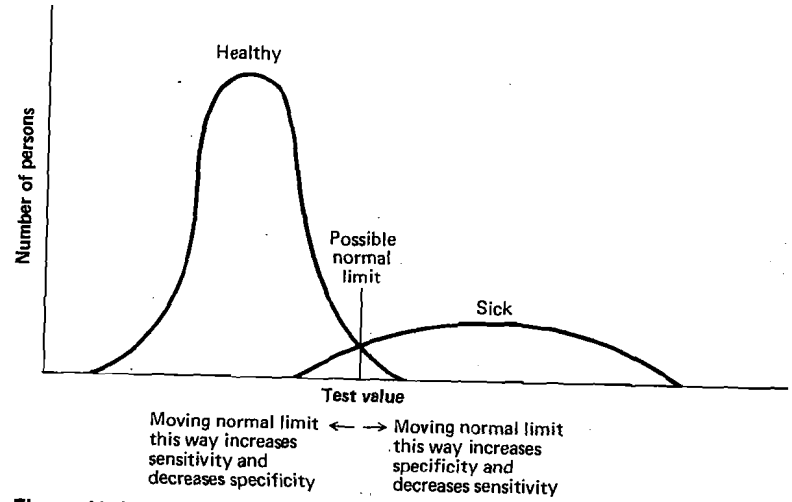

Figure 13-1 Typical example of the overlapping distributions of a test value in the healthy and the sick. Effects of shifting cutoff point on sensitivity and specificity.

factors such as blood pressure and cholesterol, it appears that within the range usually observed in this country there is no cutoff point between a safe and unsafe level. That is, the lower the level, the better off one is. There is no single level above which treatment should be given. Decisions to treat an elevated coronary risk factor must be based not only on the level of the risk factor itself but on the presence or absence of other indicators of risk. Thus, one is more apt to try to lower a serum-cholesterol level of 260 mg/100 ml if it is found in a middle-aged man who also smokes cigarettes and has a blood pressure of 150/95.

Serum cholesterol provides an excellent demonstration of the distinction between normal-usual and normal-healthy. It is not at all unusual to find a middle-aged man with a serum-cholesterol level which should hardly be considered as indicative of good health, even if the man feels well. Specifically, there is abundant epidemiologic evidence that *one-fourth* of men, i.e., those who happen to belong to the highest quartile of the cholesterol distribution, have three to four times the risk of developing clinical coronary heart disease as men in the lowest quartile. Knowing that about one in

every five men in this highest quartile will develop clinical coronary heart disease in the next 10 years hardly leads to confidence that they are normal-healthy, especially since there is mounting evidence that medical attention to their diet and living habits may reduce their high risk.

Epidemiologic studies have shown that many characteristics that have been regarded as normal because they are usual in persons who presently feel well are associated with a high probability of *future* disease. A preventive approach implies that these characteristics can no longer be regarded as consistent with good health.

Practicing Preventive Medicine in the Office or Clinic

Epidemiologic knowledge fosters the practice of preventive medicine in the medical office. Knowledge of the factors and characteristics which *cause* or *predict* the development of a disease permits identifying individuals who are at high risk of developing it. It may then be possible to prescribe measures for these patients that will prevent or at least delay the onset of the disease.

Nowadays, pediatricians are quite comfortable with the office practice of preventive medicine. Immunizations and well-baby care constitute an important phase of their work. However, in adult care the trend toward preventive medicine is only gradually taking hold. In the hope of hastening this process, the simplicity and ease of preventive care for one of the major threats to adult health, coronary heart disease, will be described.

Based on our current knowledge of coronary risk factors, identification of high-risk individuals requires little cost and effort. Age, sex, family history, and pertinent habits such as cigarette smoking and exercise are simple historical items that can be obtained by paramedical personnel or self-administered questionnaires. Likewise, height and weight or simple observation to detect obesity and a blood-pressure measurement can be done by anyone in the office with minimal training. All that remains is an electrocardiogram and the drawing of a blood specimen for measuring cholesterol and glucose. A fasting blood specimen for triglyceride might be

added, but there is some evidence suggesting that this adds little if the cholesterol is known.

Most authorities presently believe that persons with elevated levels of correctable risk factors should receive remedial therapy. As mentioned earlier, there is no safe cutoff point for each measure. The physician must form an overall impression of the patient's risk and act accordingly.

The remedial measures for the most part appear to be safe and consistent with good general health. Where appropriate, advice should be given to stop smoking, to eat less rich food in order to reduce weight and blood lipids, and to get more exercise without going to sudden extremes. Drugs that are apparently safe will lower blood pressure in most cases and will reduce lipid levels that do not respond sufficiently to diet.

The point that requires emphasis is that the detection and treatment of high coronary risk is simple and should be well within the scope of office medical practice. It would be naïve to assume that all high-risk patients will stop smoking, or eat less, or exercise more, if a doctor tells them to. Probably most will not. But some will, and it would be a shame if they were not given the opportunity and encouragement to lower their risk for a frequent, often fatal, disease.

CRITICAL READING OF THE MEDICAL LITERATURE

Most health-care professionals do not have enough time available for the careful reading and study of all the medical and scientific articles that come to their attention. It is important, however, to be able to evaluate critically reports and papers that can influence clinical decision or practice.

It is not intended, in this brief discussion, to cover all the errors and pitfalls that can occur in medical papers. Evaluating methods, observations, and interpretations in specialized fields such as surgery or biochemistry often requires knowledge and experience in the particular discipline.

An understanding of epidemiology does foster a critical approach to certain aspects of papers involving the study of populations or patient groups. The following discussion will focus primarily

on some common problems and fallacies of an epidemiologic or statistical nature. The basic principles involved should already be familiar to the reader as they have been mentioned in previous chapters.

Need for an Adequate Control Group or Basis of Comparison

Many papers report findings apparently showing the benefits of a preventive or treatment measure, based on what appear to be good results, when the measure has been used on a study group. In viewing these "good results" the reader should always ask, *Compared to what?* This initial question will usually imply others such as, *Was there a control or comparison group? Who constituted the control group? Were they similar to the treated group in all important aspects other than the treatment?* The author should have provided clear and satisfying answers to these questions in the paper. If not, there is good reason to doubt the claimed benefits.

It might be found, for example, that 95 percent of those given a certain hypnotic drug reported the next day that they slept soundly. Although, at first glance, this in itself sounds like impressive evidence for the efficacy of the drug, we must know what percentage of similar but untreated persons would report sleeping soundly on the previous night. Furthermore, to rule out a placebo effect, we need to know what percentage, given an inactive "sleeping pill," would similarly report sound sleep.

The demonstration of harmful effects also requires a basis of comparison. It may be recalled from Chap. 7 that it was not sufficient to show that a large proportion of fatally injured pedestrians have high blood-alcohol levels to incriminate alcohol as a contributor to being struck and killed by a motor vehicle. It was also necessary to demonstrate that noninjured pedestrians, otherwise similar to those killed, had, on the average, *less* alcohol in their blood.

Requirement of Denominators for Statements Comparing Risks

Statements implying that a factor involves greater or less risk of a certain outcome are often made using only "numerator" data. The

reader should "think epidemiologically" and remember that statements concerning risk should be based on *rates*, which require *denominators* as well as numerators. An example, again regarding motor-vehicle accidents, comes from a radio advertisement of a few years ago promoting the use of auto seat belts. A statement was made to the effect that 75 percent of all motor-vehicle fatalities occurred within 25 miles of home. The implication seemed to be that it was especially risky to drive on short trips close to home. However, note that motor-vehicle fatalities constitute only the numerator of a mortality rate, which needs also an appropriate denominator, such as passenger-miles. If, say, 95 percent of all passenger miles were driven within 25 miles of home, it could easily be shown that the *risk* of getting killed per passenger mile is *less* within 25 miles of home than it is farther away.

Failure to choose the appropriate denominator in drawing conclusions about risk is an easy error to fall into. One might note that the age distribution of a large series of 500 myocardial infarction cases observed at a particular institution was as shown in Table 13-1.

It would be tempting but erroneous to conclude on the basis of these data that the risk of myocardial infarction rises with age into the sixties and then falls sharply. Statements about risk at various ages must be related to the underlying population from which the cases are drawn. Incidence rates should be constructed by using the number of cases at each age as the numerator and the number of

Table 13-1 Hypothetical Age Distribution of Myocardial Infarction Cases

Age	Number of cases	Percent
20–29	10	2
30–39	40	8
40–49	75	15
50–59	125	25
60–69	175	35
70–79	50	10
80+	25	5
Total	500	100

persons at risk in the denominator. These will permit an appropriate comparison of risk at different ages. Fallacious inferences about risk, of the type illustrated here, are frequent and should be watched for.

Other Problems

A variety of special problems involving particular concepts, measurements, or study designs have been discussed in previous chapters. Examples are the possibility for spurious correlations due to uncontrolled variables, the need to distinguish statistical from biological significance (Chap. 11) and the likelihood of biased comparisons of survival when the starting point for follow-up is different in two groups (Chap. 10). Perhaps the reader's attention should again be called to the discussions in Chap. 3 on the limitations of medical observations, to the sections in Chaps. 4 through 10 concerning the conduct and interpretation of various types of studies, and to the interpretation of statistical associations as described in Chap. 11. In addition, much of the advice on conducting a study in Chap. 12 is also pertinent to evaluation the studies of others. Further discussion of problems and fallacies can be found in Ludwig and Collette (1971), Schor and Karten (1966), Hill (1971), and Huff (1954).

Although they are not solely epidemiologic concerns, some other general points deserve consideration in reading critically. These are discussed below.

Possibilities for Bias

The possibilities for biased comparisons are many. Misleading differences between groups may result from differences in the way they were selected, differences in the way data were collected from them, different follow-up durations, different criteria for judging outcome, and so on. The critical reader should try to think of these sources of bias and should note whether the author has taken them into account in his study methods or data analysis. Important potential biases should at least be mentioned in the Discussion section of a paper, if they could not be excluded.

Need for Adequate Information

It is important to determine whether the author has described his methods of selecting subjects, and of collecting and analyzing data in enough detail so that they can be evaluated and so that others can try to repeat the study or understand why their findings might differ. By close attention to these methods, the critical reader may also be able to determine whether the study was done with care or rather haphazardly.

Evidence of Objectivity

Some attempt should be made to determine whether the author appears to be objective or whether he is an advocate of a particular point of view. Is the presentation slanted toward a particular viewpoint? Would the author have published the paper if the opposite findings had been observed? Some knowledge of his previous work may be helpful in answering these questions.

One way that lack of objectivity may affect study results is through a selection process. Without intending to be misleading, an investigator may emphasize those observations which support his point of view and discard those that do not. Referring again to Fig. 11-1, page 152, which shows a moderate correlation between coronary heart disease mortality and per capita cigarette consumption in 44 states, note that the points for Utah, Arkansas, Kentucky, Indiana, and Connecticut fall along a straight line. If one wanted to show that the two variables had a nearly perfect correlation, one could prepare a graph showing these five states only. These five points would indeed present an impressive picture, if it were not mentioned that they were selected out of all the available data.

Selection for Publication

Viewing the medical literature as a whole, it is clear that positive findings are more apt to appear than negative findings. It must be remembered that positive findings may occur by chance where there is no relationship. Even when the authors are objective, chance positive findings are more apt to find their way into the literature

than truly negative findings, at least until controversy makes negative findings just as important and interesting as positive findings.

REFERENCES

Dry, T. J., *Manual of Cardiology.* (Philadelphia: W. B. Saunders, 1943) pp. 1,2.

Hill, A. B., *Principles of Medical Statistics*, 9th ed. (London: Oxford University Press, 1971), Chaps. 21–23.

Huff, D., *How to Lie with Statistics.* (New York: W. W. Norton, 1954).

Ludwig, E. G., and J. C. Collette. 1971. Some misuses of health statistics. *J. Am. Med. Assoc.*, **216**:493–499.

Price, W. H. 1963. Gall-bladder dyspepsia. *Brit. Med. J.*, **2**:138–141.

Schor, S., and I. Karten. 1966. Statistical evaluation of medical journal manuscripts. *J. Am. Med. Assoc.*, **195**:1123–1128.

Weiss, N. S. 1972. Relation of high blood pressure to headache, epistaxis, and selected other symptoms: The United States health examination survey of adults. *New Engl. J. Med.*, **287**:631–633.

Chapter 14

Epidemiology, Medical Care, and the Health of the Community

Health and disease in the community are important concerns not only of medical and public health professionals but of the general public as well. To illustrate the important role of epidemiology in community health, two types of epidemiologic investigations will be described briefly—the time-honored investigation of infectious-disease epidemics and some recent efforts to detect unsuspected environmental hazards. Then, the limited effects of medical care on community health will be discussed. Screening for disease and other methods for increasing the beneficial effects of health care on the community will then be described.

Investigation of Epidemics of Infectious Disease

Until a few decades ago, epidemiology had focused primarily on the infectious diseases, which have been the major scourges of mankind. Recently, in the more affluent nations, most infectious diseases

have been brought under reasonably good control, and the leading causes of death and disability have become the noninfectious conditions. Thus, in these areas, many epidemiologists' attention has been directed toward chronic degenerative and neoplastic diseases. Other diseases or conditions of great interest and importance include physical trauma or accidents, mental illness, and congenital defects. Additional concerns that have recently engaged the epidemiologist are studies of medical care and health services, and studies that focus on general health status irrespective of the particular diseases responsible for departures from good health.

Despite a shifting emphasis in the more affluent nations, infectious diseases remain extremely important problems in the "less developed" parts of the world. Furthermore, dangerous infectious-disease outbreaks continue to occur in industrialized nations. Even though the principal causes of many of these diseases are fairly well understood, epidemiologists, health officers, other physicians and public-health nurses are still called upon to investigate specific disease outbreaks to determine the particular conditions or factors that are responsible.

Investigation of the variety of epidemics that might occur cannot be described by a single step-by-step "cookbook" approach. However, certain principles are followed sufficiently often to deserve at least a brief summary here. The interested reader should consult Anderson et al. (1962) for more details.

The typical field investigation of an epidemic involves, first, a study of the cases. Clinical examination and appropriate laboratory tests are needed to determine or verify the diagnosis. Once the disease is identified, knowledge of the usual sources of infection and common modes of spread for that disease will help point the investigation toward the most likely explanations of the epidemic. A convenient reference book that summarizes the important information for most infectious diseases is *Control of Communicable Diseases in Man*, published by the American Public Health Association (Benenson (ed.), 1970).

In addition to verifying the diagnosis, the patients are studied further, usually by interview. Their basic characteristics such as age, sex, and occupation should be determined, as should the onset and

time course of the disease. Personal contacts at home, work, school, and other places, special events such as parties and trips, foods eaten, and exposures to other common vehicles are items that will frequently be inquired about, depending, of course, on the disease believed responsible for the outbreak. Particular emphasis should be placed on the time period when the patient was most probably infected. This period precedes the disease onset time by an interval equal to the usual incubation period for that disease.

The subsequent investigation will be guided by information gained from the known cases. For example, plotting the dates of disease onset graphically as in Figs. 5–7 (page 69), 5–8 (page 71), and 5–9 (page 72) will help determine if the epidemic had a point-source or involved person-to-person spread. Or, if the disease involves a gastrointestinal infection and several cases mention going to the same restaurant or party, further investigation of possible food contamination at the restaurant or party would be in order. Pursuing the party further, apparently well persons who also attended it might also be given appropriate laboratory tests to detect subclinical infection. Comparisons of what the infected and uninfected persons ate at the party will help determine which foods were contaminated. A good example of the simple analyses that are made to identify foods that serve as vehicles for infection is presented and discussed by Sartwell (1965).

Data analysis concerning possible causative factors for an epidemic will usually take the form either of an incidence study or a case-control study. In the former approach the incidence, or "attack" rates, of persons exposed to possible sources or vehicles are compared with those of persons not exposed. If the rates are much higher in the exposed, the source is highly suspect. In the case-control comparison, the suspect sources are those to which a much higher proportion of cases than controls were exposed.

It is hoped that investigation of the epidemic will reveal correctible problems. A major accomplishment would be the identification of infected persons who can continue to spread disease if not attended to, such as typhoid carriers working in restaurants or hospital employees with staphylococcal skin infections. The recognition of other factors leading to the spread of disease, such as

improper food-handling practices, contaminated water supplies, or segments of the population who have not received the usual vaccinations, can also lead to effective control measures.

The Detection and Evaluation of Environmental Hazards

In recent years there has been considerable concern that we are poisoning ourselves with our technology. It is well known that our land, water, and air are being polluted by such substances as industrial wastes, exhaust products from burning fuels, trace metals, chemicals, pesticides, and radioactive materials. Furthermore, the population now ingests a variety of chemicals in such forms as preservatives and medicinal drugs.

What is less clear is the extent to which these substances affect human health. Epidemiologic studies can play an important role in the quantitative determination of the risks involved.

The usual investigations have employed standard epidemiologic methods to assess the relationship between specific substances, drugs, energy sources, or occupational exposures, and particular disease outcomes of interest. Examples are the Berlin, New Hampshire prevalence study of chronic respiratory disease in relation to air pollution and smoking, described in Chap. 6, the case-control study of thromboembolic disease in relation to oral contraceptives described in Chap. 7, and the cohort study of occupational exposure to x-ray described in Chap. 8. (For further examples, see Whittenberger, 1967, and Goldsmith, 1972).

These studies involve an assessment of environmental hazards that are *already under suspicion.* The proliferation of new chemicals and energy sources to which we are exposed has led to serious concern that there are many hazards that we are not aware of. Sometimes, unsuspected hazards come to light only after considerable damage has been done. A recent dramatic example involved the drug thalidomide which, when given as a tranquilizer to pregnant women, resulted in the birth of thousands of seriously deformed babies. Other classical cases were the occurrence of retrolental fibroplasia in premature infants who received oxygen therapy, and the development of bone cancers in radium-dial painters.

As a result of concern for the unsuspected, epidemiologists

have begun to work in a new area of research, sometimes called *monitoring.* The purpose of monitoring is to detect unsuspected adverse or undesired effects of components of the environment as soon as possible after these effects appear, and thus provide an "early warning system." Because broad areas are to be covered, this type of investigation usually involves initially a search for hypotheses. Suspicious relationships can then be subjected to more intensive study.

Much of the experience to date in monitoring environmental hazards has been gained from monitoring adverse reactions to medicinal drugs (Report of the International Conference on Adverse Reactions Reporting System, 1971). Despite careful testing of drugs before they are marketed, many drug reactions do not become recognized until the drug has been extensively used by large numbers of patients. Monitoring of drug reactions will be used to illustrate some of the methods to be described.

A number of methods or systems of monitoring are available. These have been tried with varying degrees of success. They involve the assembly and analysis of data on morbid events, usually, but not always, in relation to various exposures.

Spontaneous Reporting Systems Many hypotheses come from the observation by individual physicians of patients who develop what seems to be an adverse reaction to a drug. Ordinarily, if he is sufficiently concerned, the physician might report this to the drug manufacturer or publish a letter or brief case report in a medical journal.

Programs known as spontaneous reporting systems have been established to encourage physicians to send such reports to a central agency or clearing house where they can be assembled and evaluated. Examples are the reporting programs that have been set up by the Food and Drug Administration in the United States and the Committee on Safety of Drugs in the United Kingdom.

While some useful information has been obtained at low cost from spontaneous reports, certain deficiencies are apparent. Despite promotional efforts to get physicians to respond, the number of reports submitted and the amount of information contained in each report have often been disappointing. Furthermore, it is very difficult

for the physician to determine the cause of a serious untoward event in a single patient. While it could be an adverse effect of a drug, it could also be a worsening or complication of the disease being treated, or even the result of a different disease that has developed independently. Physicians tend to recognize and interpret as drug reactions events that they are familiar with in that regard, such as skin rash following penicillin therapy or aplastic anemia following chloramphenicol; they tend not to report unsuspected relationships. Finally, the lack of any "denominator" or "population-at-risk" information makes it difficult to determine whether the reaction might be rare or relatively common.

Changes in Disease Frequency If a population is being exposed to a new environmental hazard or to increased levels of an old hazard, suspicion can be aroused by monitoring disease frequency. Populations or special subgroups may be kept under systematic surveillance to determine time trends in incidence, prevalence, or mortality from any or all diseases. A good example is the monitoring of congenital malformations in newborn infants. The prevalence of various malformations among newborn infants in several cooperating hospitals can be recorded on a monthly basis. If the occurrence of one or more malformations shows a distinct increase beyond the fluctuations usually noted due to chance or seasonal variations, then an inquiry into prenatal exposures might be initiated, much as one would investigate an epidemic of infectious disease (Hook, 1972).

Although probably less accurate than special programs to monitor disease frequency, the surveillance of vital statistics can also provide useful information about changes in disease frequency. Increases in mortality rates in communities or increases in congenital malformations reported on birth certificates can provide useful clues that something is happening in the environment.

Intensive Surveillance of Both Exposures and Disease Procedures can be established to collect extensive information concerning both exposures and disease frequency. In this way a variety of exposure/disease relationships can be explored to look for unsuspected relationships and to develop quantitative information about known relationships. An example of this type of program is the

Community Health and Environmental Surveillance System (CHESS) of the Environmental Protection Agency, in which several components of air pollution and several measures of health and disease are measured in selected communities (Riggan et al., 1972). Examples in drug-reaction monitoring are the Boston Collaborative Drug Surveillance Program (Jick et al., 1970) which collects and analyzes data about drug administration and untoward events from several hospitals, and the Kaiser-Permanente Drug-Reaction Monitoring System, which emphasizes drug use and drug reactions in outpatients (Friedman et al., 1971).

Limited Effects of Medical Care—Historical Perspective

With the impressive technical advances in medical care that have become available in recent decades, it is easy to forget that the quality and quantity of medical care have only a limited influence on community health. That medical care is not the only determinant of health is well illustrated by the observed long-term time trends in mortality from certain diseases. As will be shown, these trends appear to bear little relationship to changes in medical care.

One example is the marked decline in mortality from tuberculosis in the United States during the twentieth century (Fig. 14-1). As pointed out by Winkelstein (1972), the only treatment available at the beginning of this century was rest therapy in sanatoriums, accessible only to the wealthy. This was made available to all economic classes in the 1930's, and during the same decade, collapse therapy was introduced. Chemotherapy became widely available in the 1950's. Fig. 14-1 shows that the downward trend in mortality was clearly evident before these new therapies were widely applied. Winkelstein also cited data from a study by Terris (1948) showing that the isolation of cases in treatment facilities was probably not a major determinant of tuberculosis mortality. Thus, even though therapy and isolation of cases may have accentuated the decline shown in Fig. 14-1, other important factors must have been operating.

Other diseases have also shown major secular changes that are difficult to attribute to the benefits of medical care. For example, McKeown and Lowe (1966) presented a graphic picture of the

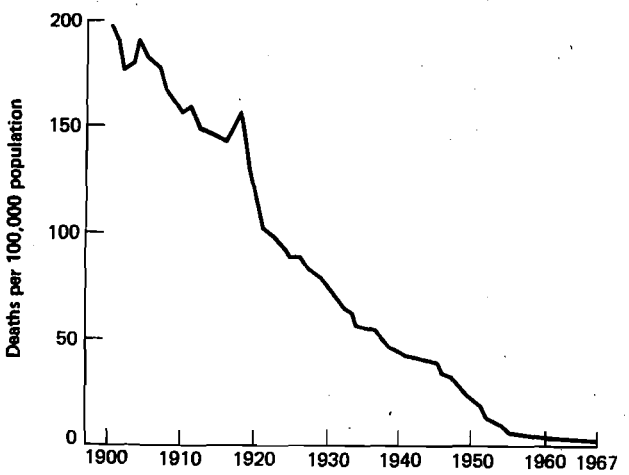

Figure 14-1 Annual age-adjusted tuberculosis death rates per 100,000 population, death registration states of the United States, 1900–1967. (Adjusted to the age distribution of the U.S. population in 1940.) *(Reproduced, by permission, from Winkelstein, 1972.)*

decline in mortality from whooping cough in English children, similar to that shown for tuberculosis in Fig. 14-1. Mortality declined rather steadily from about 1,400 deaths per million per year in 1870 to a negligible number in the 1960's. Relevant medical landmarks during this decline were the identification of the causative organism in 1906, sulfonamide and antibiotic therapy, beginning in 1939, and the general availability of immunization beginning in 1952. By the time drug therapy and immunization became available, the whooping cough mortality rate was only a small fraction of what it had been in the late nineteenth century, when it was a very important cause of death.

The rather striking time trends over the last few decades in lung cancer and stomach cancer mortality rates (see Fig. 5-10, page 75) are also largely independent of the effects of medical care. To date, medical, surgical, and radiation therapy cannot save the lives of the vast majority of victims of these two malignancies.

It is obvious from these examples and from the usual incurability of the degenerative and neoplastic diseases that constitute most of our leading causes of death and disability, that if we are to bring

these conditions under control and improve the health and longevity of the population, we cannot rely solely on increasing the availability of medical care as we know it today. More important will be to improve our understanding of the environmental and social factors which foster these diseases and, using this knowledge, to apply effective preventive measures.

Two Aspects of Disease Prevention

One approach to disease prevention is through medical care of individual patients. As an example of this approach, a simple method for preventing coronary heart disease, to be used in the clinic or physician's office, was described in Chap. 13. It involved, first, detecting individuals who are at high risk of developing the disease, and secondly, attempting to reduce their risk by changing their dietary and other habits and judiciously prescribing drugs, when indicated.

The second avenue of disease prevention does not focus on the individual. Rather, it involves large-scale social and environmental changes, such as improving housing conditions, requiring pasteurization of all milk sold commercially, or adding fluoride to community water supplies. For coronary heart disease, possible preventive approaches on this scale might include changing food processing to decrease the amount of saturated fat and cholesterol in animal food products, discouraging cigarette smoking by increased taxes or by other forms of persuasion well known to the advertising industry, or discovering the harmful agents in tobacco smoke and removing them, or banning automobiles from certain areas so as to force many people to walk or ride bicycles. These measures are listed here as examples of efforts that might be considered, not as proven practical approaches to the problem of coronary heart disease.

Because many people do not go to doctors routinely and because many others who do go either do not follow medical advice or find it extremely difficult to break pleasurable habits, one is forced into a rather pessimistic view about the impact that office preventive medicine can have on the health of the general population. Even special intensive programs to change patient behavior have not

proven to have as much long-term effect as has been hoped. For example, a variety of innovative methods have been tried to help people stop smoking cigarettes. Despite high initial success rates, follow-up one year later usually reveals that only about one-fifth of those originally treated still refrain from smoking.

The health-care professional must do what he can to help his patients. Certainly a success rate of one out of five is better than nothing. Nevertheless, since medical care often has so little impact on major health problems of the community, many believe that only large-scale social and environmental changes will be effective.

The Physician's Limited View of Disease in the Community

One reason that medical care has less influence than one might expect is that much disease never comes to the attention of medical personnel. Using prevalence survey data obtained in Great Britain and the United States, White, Williams, and Greenberg (1961) showed how illness in the community gets filtered to various physicians and institutions in the medical-care system. As shown in Fig. 14-2, of 1,000 adults in the community, 750 report one or more illnesses each month. One-third of those with illnesses, or 250, consult a physician. Only 9, or 1.2 percent of the ill are admitted to a hospital, and only 5, or 0.7 percent, are referred to another physician. Particularly striking from the viewpoint of medical education is the fact that only one of these patients is seen at a university medical center.

Not only do just a portion of the sick get seen medically, but each medical specialty and medical setting attracts a selected group of patients out of all those seen. For example, in the outpatient clinic one is especially apt to encounter patients with mild acute and chronic illness and patients with symptoms for which no organic basis can be found. In the hospital, patients are, on the average, much more seriously ill with diseases that are farther advanced. As pointed out by White et al., "Each practitioner or administrator sees a biased sample of medical care problems presented to him; rarely has any individual, specialty or institution a broad appreciation of the ecology of medical care that enables unique and frequently

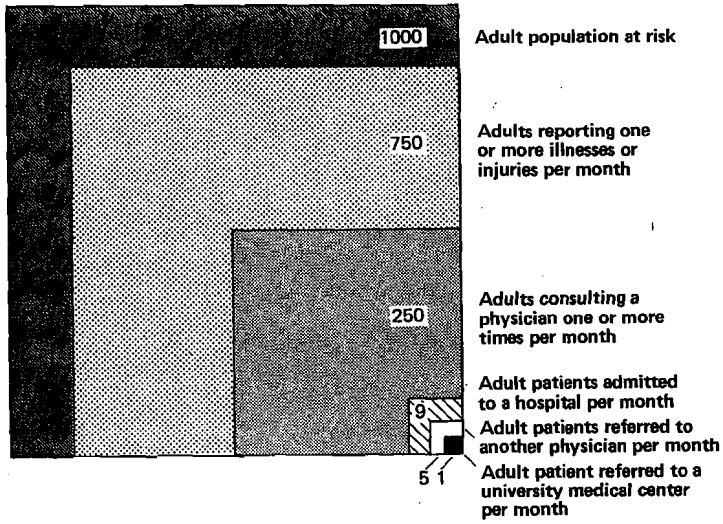

Figure 14-2 Monthly prevalence estimates of illness in the community and the roles of physicians, hospitals, and University Medical Centers in the provision of medical care (adults 16 years of age and over). *(Reproduced, by permission, from White, Williams, and Greenberg, 1961.)*

isolated contributions to be seen in relation to those of others and to the over-all needs of the community."

Efforts to Bring More Disease to Medical Attention: Screening

In the hope of increasing the impact of current medical knowledge and technology on disease in the community, medical and public health facilities have established *screening* programs to detect persons with early, mild, and asymptomatic disease.

As stated by Thorner and Remein (1961), "The basic purpose of screening for disease detection is to separate from a large group of apparently well persons those who have a high probability of having the disease under study, so that they may be given a diagnostic workup and, if diseased, brought to treatment." Since screening tests are designed to be applicable to large population groups, they

must be simple, rapid and inexpensive. As a result, they are generally less accurate and definitive than the examinations and tests used by physicians to arrive at a final diagnosis.

Screening is carried out in the belief that the detection of disease in an early or asymptomatic stage will lead to appropriate treatment which, in turn, will lead to less disability and/or mortality from the disease. After an initial period of enthusiasm in some quarters, considerable skepticism developed concerning the benefits of screening, based on doubts as to whether it would really lead to favorable modifications in the course of disease.

Critics pointed out that many persons with diseases discovered by screening did not receive adequate or appropriate treatment afterward. Or, even if the accepted treatment is given for some diseases detected by screening, such as mild maturity-onset diabetes mellitus, it has not been shown that persons so treated live happier, healthier, or longer lives. Also, persons correctly or incorrectly labeled as having a disease would be caused considerable worry and anxiety, often to no good purpose. Furthermore, careful analysis showed that apparently good results of screening could be misleading, due to self-selection for screening of those persons who take better care of themselves, or due to fallacies such as that pointed out in Chap. 10, page 148, wherein persons would *seem* to survive longer merely because the diagnosis was made earlier. Even ethical questions have been raised (McKeown, 1968) since in contrast to traditional medical care which is sought by the patient, disease detection by screening is initiated by medical or public health professionals, who, therefore, are under special obligation to make sure that screening does more good than harm.

Thus, evaluation of screening programs requires carefully controlled experimental studies in which relevant disease outcomes are measured in a group receiving screening and compared to outcomes measured similarly in a comparable unscreened group. Any benefits found for screening programs should be compared to costs involved (McKeown and Knox, 1968; Wilson, 1968; Cochrane and Holland, 1971). Although it is more an evaluation of periodic comprehensive health checkups than of screening per se, the Kaiser-Permanente Multiphasic Health Checkup Evaluation Study, described in Chap. 9, illustrates the approach that is needed if

screening programs are to prove their merits. Other examples of well-controlled evaluations of screening or disease-detection methods are the study of breast cancer screening by the Health Insurance Plan of New York (Shapiro, Strax, and Venet, 1971) and the study of lung cancer screening by Brett (1971).

A generally accepted principle is that screening should only be done if it can be integrated with the medical-care program where it is carried out. In practice this means not only that adequate treatment, care, and follow-up be available for those who screen positive, but that the screening test results must be acceptable to the practicing physicians in the area. The characteristics of screening tests that relate to accuracy and acceptability will be discussed briefly.

Sensitivity and specificity, two measures of the accuracy of diagnostic tests, were defined in Chap. 13, page 192. These measures are also important features of screening tests. The relationship of a screening test to the final accurate diagnosis is conveniently shown in a fourfold table (see Table 14-1). In the table, *a* represents persons with the disease who are correctly labeled by the screening test. Persons denoted by *b* are *false positives*, since the test is positive but they do not have the disease. The letter *d* denotes persons free of disease who are correctly labeled by the test. The letter *c* represents *false negatives*, persons with the disease for whom the test is negative.

Sensitivity, the proportion of true positives that are labeled as positive, is thus $a/(a + c)$. Specificity, the proportion of true negatives that are labeled as negative is $d/(b + d)$. Both of these measures are important, since the test should detect as much disease as possible while avoiding false labeling. False negatives,

Table 14-1 Relationship of Screening-Test Results to the Final Accurate Diagnosis

Screening test	Final diagnosis: Disease present	Final diagnosis: Disease absent	Total
Positive	a	b	$a + b$
Negative	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

persons with undetected disease, may be deprived of valuable therapy. False positives, persons incorrectly labeled as diseased, are subject to needless worry and expensive diagnostic evaluations until their freedom from the disease is established. With all quantitative screening tests, the level above or below which a person is called positive will affect the sensitivity and specificity. Modifying this cutoff level to improve one of these characteristics will adversely affect the other.

Physicians evaluating patients who have been screened are especially sensitive to another measure, $a/(a + b)$, the proportion of positive tests that are true positives. Since physicians are usually asked to evaluate only the positive screenees, they understandably become irritated and critical of the screening program when most of their follow-up diagnostic evaluations turn out to be negative.

If the disease is infrequent in the population—and most chronic diseases are—even a screening test with a high degree of specificity will yield positives of which a large percentage turn out to be false. Thorner and Remein (1961) showed an example of a population of 10,000 with an assumed prevalence of diabetes mellitus of 1.5 percent, screened with a random blood glucose (not drawn at any particular time in relation to eating). Using a cutoff point of 130 mg percent, the test has been shown previously to have a sensitivity of 44.3 percent and a high specificity of 99.0 percent. The results are shown in Table 14-2. Note, that of the 164 positives, 98, or 60 percent, turn out to be false positives.

If, in order to decrease the number of false positives, the screening level is raised to 180 mg percent, the specificity will now

Table 14-2 Results of Screening for Diabetes Mellitus in a Population of 10,000*

Screening test	Final Diagnosis		Total
	Diabetic	Not diabetic	
Positive	66	98	164
Negative	84	9,752	9,836
Total	150	9,850	10,000

*In this population the disease prevalence is 1.5%, and the sensitivity and specificity of the test are 44.3% and 99.0%, respectively.

Source: Data from Thorner and Remein (1961).

be 99.8 percent. The test will now yield only 54 positives, of whom only 20, or 37 percent, are false positive. However, there is a marked decrease in sensitivity. Only 34 of the 150 diabetics will be detected.

If the disease prevalence is higher, a larger proportion of positives will be true positives. One strategy for increasing the prevalence of disease in the population screened is to restrict screening to high-risk individuals. For example, screening for diabetes by measuring blood sugar may be carried out only among persons who are obese or who have a family history of the disease.

Broadening the Concept of Screening

Although formal screening programs were initially directed primarily at the early detection of single specific diseases, the screening concept has expanded in recent years to encompass *screening for high risk* and *multiphasic screening.*

Screening for High Risk As more emphasis is being placed on disease prevention, community programs for disease control may well include screening programs to detect persons at high risk of developing disease. In this way, preventive measures can be applied before the disease occurs. For example, pilot programs are now underway in industry and communities to identify persons with coronary risk factors such as high serum-cholesterol and blood-pressure levels, so that myocardial infarction and other manifestations of coronary heart disease can be prevented. The long-term effects of these programs need to be evaluated.

Before setting up such a program it is necessary, as with screening for frank disease, to make sure that suitable care and follow-up will be available for positive screenees. That is, the screening must fit in with the local medical care program so that something more than patient anxiety will result.

Multiphasic Screening It is more efficient to screen for a variety of diseases at one time than to carry out separate screening programs for single diseases. Fostered by the development of automated testing procedures, *multiphasic* screening programs are becoming widespread.

Multiphasic screening or multiphasic health testing is being viewed increasingly as having greater utility than just in the detec-

tion of asymptomatic disease (Thorner, 1969). It has been shown to be an efficient and economical component of periodic health checkups for patients both with and without known disease. Used in this way, multiphasic screening of high quality appears to be acceptable to both physicians and patients, and it conserves physician time and other medical-care resources (Collen, 1971).

Multiphasic health testing is also seen now as an important component of a new mode of entry of patients into medical care. With the trend toward prepayment or government payment for medical care, the traditional economic barrier, the fee for service, is disappearing. To prevent a resultant overloading of the medical-care system and to assure appropriate allocation of physician time to the care of the sick, Garfield (1970) has proposed a new system of organization for medical care. He suggested that patients not acutely ill enter the system in a way that would utilize multiphasic health testing to help determine the nature of the problem and the appropriate facility to which the patient should be referred.

Evaluating a Changing Health-Care System

We are living in a period of great change in health care. Spurred by technological and socioeconomic advances, many old methods are being questioned or discarded and new approaches are being introduced.

Innovations can and should be evaluated by well-controlled experiments, whenever possible. Where particular circumstances or meager resources prohibit rigorous experiments, less-formal evaluations such as before/after comparisons can be conducted. However, careful attention should then be paid to extraneous influences and biasing factors that may affect the apparent outcomes of the innovation.

A few of the health care issues of current interest are modes of payment for services, the use of nurses and other paramedical personnel for tasks traditionally performed by physicians, the content and frequency of health checkups and community screening programs, control of drug overuse and abuse, and provision of optimal care for persons living in inner cities and remote rural areas. Defining and describing these problems, and identifying and evaluating possible solutions all involve studies of health-related characteristics, events, and outcomes in groups of people.

Whether these studies are labeled as "epidemiology," or "medical-care research," or "health-services research" makes little difference. What is important is that we be guided by careful observations and wise judgment to make necessary improvements while preserving the many good methods and approaches that we now have.

REFERENCES

Anderson, G. W., M. G. Arnstein, and M. R. Lester, *Communicable Disease Control: A Volume for the Public Health Worker*, 4th ed. (New York: Macmillan, 1962) Chap. 10.

Benenson, A. S. (Ed.), *Control of Communicable Diseases in Man.* 11th ed. (New York: American Public Health Association, 1970).

Brett, G. Z. 1968. The value of lung cancer detection by six monthly chest radiographs. *Thorax*, **23**:414–420.

Cochrane, A. L., and W. W. Holland. 1971. Validation of screening procedures. *Brit. Med. Bull.*, **27**:3–8.

Collen, M. F. 1971. Guidelines for multiphasic health checkups. *Arch. Intern. Med.*, **127**:99–100.

Friedman, G. D., M. F. Collen, L. E. Harris, E. E. Van Brunt, and L. S. Davis. 1971. Experience in monitoring drug reactions in outpatients: The Kaiser-Permanente Drug Monitoring System. *J. Am. Med. Assoc.*, **217**:567–572.

Garfield, S. R. 1970. The delivery of medical care. *Sci. Am.*, **222**:15–23.

Goldsmith, J. R., Statistical Problems and Strategies in Environmental Epidemiology. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. vol. VI Effects of Pollution on Health.* (Berkeley: University of California Press, 1972), pp. 1–28.

Hook, E. B., Monitoring Human Birth Defects, Methods and Strategies. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. vol. VI Effects of Pollution on Health.* (Berkeley: University of California Press, 1972), pp. 355–366.

Jick, H:, O. S. Miettinen, S. Shapiro, G. P. Lewis, V. Siskind, and D. Slone. 1970. Comprehensive drug surveillance. *J. Am. Med. Assoc.*, **213**:1455–1460.

McKeown, T.: Validation of Screening Procedures, in *Screening in Medicine: Reviewing the Evidence: A Collection of Essays.* (London: Oxford University Press, 1968), pp. 1–13.

McKeown, T., and E. G. Knox, The Framework Required for Validation of Prescriptive Screening, in *Screening in Medicine: Reviewing the Evidence: A Collection of Essays.* (London: Oxford University Press, 1968), pp. 159–173.

McKeown, T., and C. R. Lowe: *An Introduction to Social Medicine.* (Philadelphia: F. A. Davis, 1966), pp. 86–87.

Report of the International Conference on Adverse Reactions Reporting Systems. National Academy of Sciences, Washington, D.C. 1971.

Riggan, W. B., D. I. Hammer, J. F. Finklea, V. Hasselblad, C. R. Sharp, R. M. Burton, and C. M. Shy, CHESS, A Community Health and Environmental Surveillance System. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. vol. VI Effects of Pollution on Health.* (Berkeley: University of California Press, 1972), pp. 111–123.

Sartwell, P. E. (ed.), *Maxcy-Rosenow Preventive Medicine and Public Health*, 9th ed. (New York: Appleton-Century-Crofts, 1965), pp. 16–19.

Shapiro, S., P. Strax, and L. Venet. 1971. Periodic breast cancer screening in reducing mortality from breast cancer. *J. Am. Med. Assoc.*, **215**:1777–1785.

Terris, M. 1948. Relation of economic status to tuberculosis mortality by age and sex. *Am. J. Public Health*, **38**:1061–1070.

Thorner, R. M. 1969. Whither multiphasic screening? *New Engl. J. Med.*, **280**:1037–1042.

Thorner, R. M. and Q. R. Remein, *Principles and Procedures in the Evaluation of Screening for Disease.* U.S. Department of Health, Education, and Welfare. Public Health Monograph no. 67, 1961.

White, K. L., T. F. Williams, B. G. Greenberg. 1961. The ecology of medical care. *New Engl. J. Med.*, **265**:885–892.

Whittenberger, J. L., The physical and chemical environment, in *Preventive Medicine*, edited by D. W. Clark, B. MacMahon, (Boston: Little, Brown, 1967), pp. 630–638.

Wilson, J. M. G. November 1968.: The evaluation of the worth of early disease detection. *J. Roy. Coll. Gen. Pract., Suppl.*, **2**:48–57.

Winkelstein, W., Jr. 1972. Epidemiological considerations underlying the allocation of health and disease care resources. *Intern. J. Epidemiology*, **1**:69–74.

Index