coronary profile: 12-year follow-up in the Framingham study. *J. Occup. Med.*, **9**:611–619.

Kannel, W. B., T. R. Dawber, A. Kagan, N. Revotskie, and J. Stokes. 1961. Factors of risk in the development of coronary heart disease: six-year follow-up experience: The Framingham study. *Ann. Intern. Med.*, **55**:33–50.

Kannel, W. B., E. J. LeBauer, T. R. Dawber, and P. M. McNamara. 1967. Relation of body weight to development of coronary heart disease: The Framingham study. *Circulation*, **35**:734–744.

MacMahon, B., Cancer. Chap. 24 in *Preventive Medicine*, edited by D. W. Clark and B. MacMahon, (Boston: Little, Brown, 1967), pp. 423–426.

Whittenberger, J. L., The physical and chemical environment. Chap. 34 in *Preventive Medicine*, edited by D. W. Clark and B. MacMahon, (Boston: Little, Brown, 1967), pp. 630–638.

Seltser, R., and P. E. Sartwell. 1965. The influence of occupational exposure to radiation on the mortality of American radiologists and other medical specialists. *Am. J. Epidemiology*, **81**:2–22.

Chapter 9

Experimental Studies

Experimental studies resemble incidence studies in that they require follow-up of the subjects to determine outcome. However, the essential distinguishing feature of experiments is that they involve some *action* or *manipulation* or *intervention* on the part of the investigators; that is, something is done to at least some of the study subjects. This contrasts with incidence and other observational studies, where the investigators take no action, but only observe.

Experiments are believed to be the best test of a cause-and-effect relationship. Something is done to an *experimental group* and the observed outcome is presumed to be the effect of that action, provided that the same outcome did not occur in an equivalent *control group* that was not acted upon. A cause-and-effect relationship can also be demonstrated by *removing* or *reducing* the alleged causal factor in the experimental group and showing a disappearance or reduction in the effect, while no change is observed in the control group.

The latter approach is especially relevant to epidemiologic experiments in preventive medicine (Hutchison, 1967). If a factor is removed or reduced and the disease incidence declines as a result, the factor is, for practical purposes, a causal one.

Although great value is placed on experimental evidence, experimental studies are often exceedingly difficult to carry out. In addition, they raise some ethical issues which must be considered.

Ethical Problems

In observational studies, the investigator's chief ethical problem, aside from the need for objectivity and conscientious work, is to maintain the confidentiality of his records about each person studied. Harm might come to an individual if some of his characteristics, recorded in confidence for medical or scientific purposes, were made available to others, or were communicated to the individual, himself, in an inappropriate manner. In the main, though, the observational epidemiologist is a passive observer of nature with few ethical problems.

The experimentalist's ethical position is quite different, since he takes it upon himself to do something to people. He must have good reason to believe that what he proposes to do has an excellent chance of helping them. On the other hand, he must also have ample doubt about the value of what is to be done, compared to doing nothing or doing what had been done in the past. Otherwise he could not, in good conscience, subject the control group to no action or to the traditional action.

Thus, medical experiments can only be carried out in a situation of uncertainty. Unfortunately, some potential investigators are so convinced as to the benefits of a treatment or preventive measure, that they are unwilling to carry out a controlled experimental test of its effects. Their *feeling* of certainty, even if based on inadequate evidence, makes them reluctant to withhold the treatment from a control group. Similarly, the unreasonable skeptic, convinced of the value of either the traditional treatment or doing nothing, may be unwilling to try new methods on an experimental basis. Both types of "believers" should realize that the failure to carry out a controlled experiment, when it is needed and feasible, is also unethical (Hill, 1971).

Sensitivity to the ethical aspects of human experimentation has resulted in the formation of committees in universities and other research institutions to review and approve all proposed studies of human subjects. It is now commonly believed that whenever possible, the potential subject should share in the decision as to whether he or she should participate in the study. This decision should be made with adequate understanding of the potential risks and benefits involved. Accordingly, informed consent is generally required from experimental subjects or from appropriate relatives or guardians.

How Experiments Are Carried Out

Experimental epidemiology is concerned primarily with testing the efficacy of measures to *prevent* disease. The preventive measure to be tested is applied to a group of persons. The incidence of the disease or disease-related outcome, such as disability, is measured in this experimental, or treated, group.

In order for the experiment to be informative, it must be controlled; that is, the outcome must be compared to some standard to determine whether any benefit has resulted. The standard may be the outcome in another similar group who do not receive the preventive measure. This control group may, instead, receive either no preventive measure or whatever is currently being applied.

Experiments may involve comparisons among several groups. For example, different amounts or dosages of the treatment may be tested. Or, there may be two or more aspects or elements in a preventive program. In this case, each experimental group may receive a different element or combination of elements. Experiments may even be designed in a more complex fashion so that each group receives a variety of treatments in sequence, possibly including periods of time with no treatment (Smart, 1970).

Randomized Control Groups The traditional and most accepted means of defining the treated and control groups is to identify one large group of all study subjects and then divide them randomly into two or more groups. If only chance determines who gets into one group or another, then the usual tests of statistical significance can be applied, to see whether chance could have

produced the observed outcome. Random assignment to groups should be done *after* the subjects are shown to be qualified and willing to participate. This will minimize subsequent losses from one or more groups.

If it is crucial that the treated and control groups be equivalent with regard to certain characteristics that might affect the outcome, the entire study population can be divided, or stratified, into subgroups and each subgroup can then be randomly divided into treated and control subjects. For example, stratification into age subgroups can be accomplished to assure that the treated and control groups have similar age distributions.

If after randomization has taken place, the experimenter would like to be sure that some nonstratified crucial characteristic is similar in the treated and control groups, he should examine the distribution of this characteristic in the two groups. If crucial characteristics differ appreciably, then the experimenter had bad luck in the randomization process. Randomization may have to be repeated, or if not possible, the results of the experiment will have to be analyzed in a way that takes into account the differences in these important characteristics. Appropriate analytic methods are discussed in Chap. 11.

Nonrandom Control Groups Randomized control groups are not always available for epidemiologic experiments. The reason may be economic. Funds may not be adequate for careful follow-up of both a treated and control group of adequate size. Or, the extra assurance that can be provided by this more ideal method may be judged to be not worth the cost involved. Also, there may not be enough subjects available for the two groups.

Even if there are enough subjects and enough money, randomization into subgroups may be impossible or may fail in actual practice. Randomization is impossible if the preventive measure can be applied only to the entire population, as when something is added to the water supply of a total community. Or, learning of the preventive measure through conversations with members of the treated group or through publicity campaigns, the control group may adopt the preventive measure to almost the same extent as does the treated group.

If randomized control groups are not used, alternative standards of comparison are available. A comparison group may be selected from persons known to be similar to the experimental group with respect to several pertinent characteristics such as age, sex, occupation, and social class. Or, if the preventive program is applied to an entire community, a similar untreated community may be used as a control.

Another approach is to have the experimental group serve as its own control. That is, a before-after comparison is made, in which there is a baseline period of observation on the experimental group before any preventive program is applied. The disease experience during this period can be compared with what happens after the program is put into effect.

Even when a separate comparison group is used, a baseline observation period is helpful. If systematic differences between the groups are noted during the baseline period, these can be taken into account in comparing the groups after the preventive measure is applied.

Possible biases or underlying group differences should always be searched for when nonrandom control groups are used. Having a group serve as its own control seems especially attractive, since this appears to eliminate virtually all group differences. However, the control and experimental observations are made during different time periods. Thus, there is the real danger that with the passage of time, other things have happened to the study group leading to the appearance of benefit from the preventive measure when none exists, or conversely, masking true benefits. Rapid changes in diagnostic and treatment methods or even in ways of life are the order of the day; these may result in real or apparent changes in disease incidence that have nothing to do with preventive methods being tested.

Subject Cooperation Many preventive measures require the cooperation or active participation of the study subjects. Experimental evaluations of these measures must take into account the failure of many subjects to cooperate. Even after initially agreeing to participate, persons drop out of the study for a variety of reasons. Also, in the treated group there will be those who take none or only

part of the treatment. Similarly, in the control group there may be some who openly or surreptitiously obtain the treatment on their own.

Study of outcomes should not be limited to the cooperators in each group since they represent a self-selected subgroup, often characterized by higher educational level, higher socioeconomic status, more concern about health and better health habits. Furthermore, if the preventive measure is eventually adopted, it will be applied in the "real world," which also has its full share of noncooperators.

Thus, the most important comparison to be made is of the *entire* study group versus the *entire* control group. This will provide the best estimate of the overall benefit to be obtained from the preventive measure if it is put into practice.

Blind Experiments If possible, experimental subjects should be kept unaware of whether they are treated or control subjects. Then, their own prejudices or enthusiasms will not result in behavior that promotes or inhibits the recognition of disease outcomes. Often, however, the nature of the treatment makes it impossible to keep the subjects "blind" to their assignment to treated or control groups.

More important is that the *assessment* of outcome be blind. Whenever possible, the physicians or others who determine whether the disease outcome has occurred should be unaware of whether the individual is a treated or control subject. The use of objective tests and criteria for diagnosis will help prevent any bias in favor of the treated or control group.

Even when experiments are designed to be blind, the subjects or their evaluators often become aware of their status. If drugs are involved in the treatment, characteristic side effects may reveal their identity. Also, unbeknown to the investigator, medical personnel involved in the care of the subjects may have access to the code or other information which identifies treated and control groups.

Thus, blind experiments are often desired but less often achieved. As for any type of study, careful evaluation of methods and results for possible bias is necessary.

The term "double-blind" is frequently encountered. Some authors use it to refer to experiments where both the assignment to treatment or control group and the assessment of results are blind. Others use it to refer to experiments in which neither the patient nor the physician knows whether the patient is in the experimental or control group.

Sample Size Considerations and Sequential Analysis Statistical methods are available for determining in *advance* how large the treatment and control groups must be, to obtain answers of the desired precision (Ipsen and Feigl, 1970). In general, the more subjects, the greater assurance that the results of the experiment are accurate and not subject to chance variation.

The desirability of having large numbers of subjects is counterbalanced by practical considerations of cost and difficulty. Ethics also enter into decisions about sample size. The more subjects included, the more who will have received the inferior treatment, if either the experimental or control regimen proves to be better.

Sometimes subjects are brought into an experiment over a relatively long period of time rather than all at once. The results for the subjects who started early may be available before the experiment is completed as planned. It is tempting to peek at early results for a few subjects and end the experiment if a difference between experimental and control groups is apparent. Unfortunately, these preliminary findings will not have the accuracy that was originally planned and agreed upon for the experiment. Stopping the experiment at this point may seem economically or ethically justified, but unless the differences noted are striking and compelling, the investigators may later regret reaching a conclusion on the basis of incomplete data. On the other hand, treatment-control differences may be much greater than originally expected, and therefore accurately demonstrable on a small number of subjects. The investigators would certainly not wish to continue the experiment, if they could be sure that this were the case.

Sequential analysis is a relatively new statistical method which allows an experiment to be ended as soon as an answer of the desired precision is obtained. The result of the comparison of each pair of subjects, one treated and one control, is looked at as soon as it becomes available and is added to all previous results. A criterion

for deciding in favor of either the experimental or control treatment is specified in advance with the desired degree of accuracy. The comparison of a relatively small number of pairs may show sufficient differences to permit the decision to be reached. If not, the results for each additional pair are added as soon as they become available until the decision criterion is met, or until it becomes apparent that there is no appreciable difference. As soon as any conclusion is reached, the experiment is stopped. The use of sequential analysis in medical experiments is described further by Armitage (1960) and Smart (1970).

Example 1. Controlled Field Trials of Poliomyelitis Vaccine

The first poliomyelitis vaccine that was widely used in the United States was the injectable vaccine containing inactivated virus, developed by Dr. Jonas Salk. By 1953, evidence had accumulated that this vaccine could be safely administered to man and that it stimulated the production of antibody that protected against the three known types of poliomyelitis virus. What was needed next was an experimental trial of the vaccine to demonstrate whether it was safe and effective when put into general use.

A large-scale cooperative field trial was undertaken in 1954, coordinated by the Poliomyelitis Vaccine Evaluation Center at the University of Michigan (Francis et al., 1955). Through the cooperation of state and local health authorities, over 200 areas participated. These were selected partly because they had experienced higher than average poliomyelitis incidence rates in previous years.

The initial plan was to inoculate school children in the second grade and observe the first- and third-graders as a control group. Although this would not permit a blind assessment of outcome, many states had agreed to participate on this basis, and this procedure was carried out in 127 counties or towns in 33 states (called "observed areas"). Eleven states were willing to cooperate in a blind experiment with a randomized control group. In the 84 counties and towns in this latter group (called "placebo areas"), participating children in the first through third grades would all receive a series of three injections, but half would receive the vaccine and half would receive an inactive *placebo*, or *dummy*.

All children in the first through third grades of the participating schools were first identified by means of a "registration form" on which was also recorded birth date, sex, race, and previous history of poliomyelitis or disability. Each child was to give a "participation request" form to his parents. This form described the observed or placebo study and provided space for the parent to sign a request that his child participate in the study. A vaccination record form was used to record all inoculations given to each participant.

Unique identification of each child on all the forms, plus cross-checking and editing of the information was carried out to ensure a high degree of accuracy. In this study there were 200,745 vaccinated and 201,229 receiving placebo among the 1,829,916 first- to third-grade children in the placebo areas, and 221,998 vaccinated second-graders and 725,173 first- and third-grade controls among the 1,080,680 first- to third-graders in the observed areas.

The vaccination phase took place between April 26, 1954 and June 15, 1954. Participating children in each classroom received vaccine or placebo from numbered vials in such a way that all three injections would be of the same material. In the placebo areas, there were vaccinated and placebo children in virtually every class. The vial code numbers could be interpreted as representing vaccine or placebo only at the Evaluation Center. Pre- and post-inoculation blood specimens were obtained from a sample of children to assess antibody response.

During follow-up, through the rest of the year, uniform procedures were instituted to detect and investigate all suspected cases of poliomyelitis among first- through third-grade children, regardless of their participation or vaccination status. The Evaluation Center was notified of all suspected cases plus all deaths from any cause. Each local health department arranged for the complete investigation of each case. The data collected included (1) a complete clinical report including history, physical examination, and spinal fluid findings; (2) laboratory specimens, including stool and blood samples for viral and antibody studies; (3) examinations by a physical therapist to classify the patient according to physical disability; and (4) autopsies, when obtainable for fatal cases.

Checking systems plus a good deal of correspondence with physicians and other persons involved were required to make

certain that the data collected were complete. By December 31, 1954 290 case records of the total of 1,103 reported were still incomplete. A campaign of telegrams, telephone calls, letters, and field visits reduced the number of incomplete reports to 78 by the end of January, but the last delinquent report was not received until March 9, 1955.

Criteria were drawn up for interpreting the laboratory and clinical findings, and on the basis of these, the investigated cases were classified as either "not polio," "doubtful polio," "nonparalytic polio," or "paralytic polio." Paralytic cases were further divided into spinal, bulbar, bulbospinal, and fatal. These decisions were all made without knowledge of the vaccination status of the children.

The experiment clearly established the benefits of the vaccines. In the placebo areas the incidence of poliomyelitis was less than half as great in those who were vaccinated (28 per 100,000) as in those who were given placebo (71 per 100,000). Similarly, in the observed areas the incidence was 25 per 100,000 in the vaccinated second-graders and 54 per 100,000 in the first- and third-grade controls. These differences were highly significant statistically. The protection appeared to be only against paralytic poliomyelitis, since there were no appreciable differences between vaccinated and controls in the incidence of nonparalytic disease.

Supporting evidence for the vaccine's effectiveness was obtained from the antibody studies. Furthermore, cases occurring among the vaccinated tended to occur in children who received vaccine which was independently judged less effective, on the basis of antigenic response. Other detailed analyses revealed that the vaccine conferred greater protection against more severe forms of paralysis and that older children appeared to benefit more than younger ones.

No ill effects of the vaccine could be demonstrated. School absenteeism for 6 weeks after the inoculations did not differ significantly among the vaccinated, placebo, and noninoculated populations. Nor was there any difference in the occurrence of rashes or other allergic manifestations, which were very rare despite the presence of small amounts of penicillin in the vaccine and placebo. Other symptoms and illnesses at the time of the injection series were quite unusual and occurred no more often in the vaccinated than in the placebo group. The minute quantities of kidney protein in the vaccine caused some concern about possible side effects on the kidney, but none could be demonstrated in the study, nor could any deaths be reasonably attributed to the vaccine.

This study represents a major achievement in experimental epidemiology. The low incidence of poliomyelitis required that a very large population be studied to provide adequate cases to reliably demonstrate the vaccine's effectiveness. Coordinating a large-scale field trial of this nature is a difficult undertaking. This summary has emphasized study design and data collection efforts, but major problems of a logistical nature should not be forgotten. For example, hundreds of thousands of children all over the country had to be supplied with the right vaccines at the right times, and thousands of blood specimens had to be drawn and transported to 28 different laboratories.

Example 2. Fluoride and Tooth Decay

Experimental studies to test the effects of adding fluorides to community water supplies were begun around 1945. The expectation that raising the fluoride concentration of drinking water to one part per million would safely lower the incidence of tooth decay was based on a number of previous observational studies. These studies had demonstrated that ingestion of water containing large amounts of fluorides during the years of tooth enamel calcification resulted in discoloration and even pitting of the teeth. However, these "mottled" teeth appeared to be quite resistant to decay. Comparisons of dental status in communities with differing fluoride concentrations in their drinking water showed that where the level was about one part per million, the decay rates were relatively low and no disfiguring mottling of the enamel was apparent.

On the basis of these findings the water supply of certain low-fluoride communities was treated on an experimental basis to bring the fluoride concentration up to the desired one-part-per-million concentration. Since randomized control groups could not be obtained for these studies, the experiment was controlled by concurrently measuring dental health status in similar but untreated low-fluoride communities. Furthermore, the dental health of children in the treated communities was assessed before the addition of

fluoride, to provide a before-after comparison. Still another comparison was made of each treated community with Aurora, Illinois, where the naturally occurring fluoride concentration in water was 1.2 parts per million and relatively little tooth decay was observed. One of these investigations, the Newburgh-Kingston Caries-Fluorine Study (Dean, 1956, Hilleboe, 1956, Schlesinger et al., 1956, Ast et al., 1956) will be described here.

The cities studied, Newburgh and Kingston, New York are located on the Hudson River about 35 miles apart. Each had a population of about 30,000. Newburgh agreed to serve as the treated community, and beginning May 2, 1945, sodium fluoride was added to its drinking water to raise the fluoride content from about 0.1 part per million to 1.0–1.2 parts per million. Kingston agreed to serve as the control community, and its water supply with a fluoride concentration of about 0.1 part per million was left unchanged.

During the year prior to adding fluoride, baseline dental examinations were carried out on the public and parochial school children, ages 6–12, in both communities. Baseline pediatric examinations were performed on smaller samples. Kingston and Newburgh children were, at first, similar regarding both general health and the prevalence of tooth decay.

Periodic assessments of both dental and other health measures were made subsequently. Although the caries experience in Kingston children remained relatively stable, a continuing improvement was noted in Newburgh.

A final evaluation was carried out after the experiment had gone on for 10 years. Over 2,000 children, ages 6–16 were given dental examinations in each community. They were selected by taking every second school child who was present on the day of the examination. Although the clinical dental examinations were not conducted in a blind fashion, x-rays were taken and were randomized at the state health department so that the interpreters would not know whether they were reading Kingston or Newburgh films.

The data analysis was carried out for separate age groups. The Newburgh subjects, ages 6–9, had used fluoridated water all their lives. The older age groups had been exposed to fluoridation starting at later periods in their dental development, and thus might be expected to show less benefit.

The efficacy of fluoridated water in preventing dental decay was clearly shown in this experiment. One of the indexes of the prevalence of tooth decay was the number of decayed, missing, or filled (DMF) permanent teeth per 100 erupted permanent teeth. For the 6–9-year-olds, this measure was 23.1 in Kingston and 10.0 in Newburgh, a relative reduction of 57 percent of the Kingston rate. The reduction in Newburgh was present in all age groups but was relatively less in older children. Thus the DMF rates in 16-year-olds were 58.9 in Kingston and 34.8 in Newburgh, a relative reduction of 41 percent of the Kingston rate. The Kingston-Newburgh differences were found in both the clinical and x-ray examinations.

Dental-caries prevalence rates in Newburgh and other communities with experimental water fluoridation programs were reduced to levels very similar to those noted in Aurora, Illinois. Thus, artifically fluoridated water was also shown to have the same benefit as observed for the naturally occurring fluoride.

Adverse effects of fluoridation were also looked for. There were no instances of disfiguring dental fluorosis or mottling. About 18 percent of the Newburgh children were found to have questionable or mild fluorosis when examined by an expert trained in detecting the effects of fluoride. The mild changes noted would have been hardly noticeable to the average dentist. On the other hand, 19 percent of children in Kingston had nonfluoride opacities or circular patches in the enamel which would have been obvious even to the untrained eye. These were found in only 8 percent of Newburgh children.

The medical examinations, x-ray estimates of bone maturation, measures of growth and development, eye and ear tests, blood counts, and quantitative studies of urinary excretion of albumin, red blood cells, and casts, all revealed no significant differences between Kingston and Newburgh children. Vital statistics data showed no consistent differences between the two communities in cancer and cardiovascular-renal death rates or in infant mortality, maternal mortality, or stillbirth rates.

These community studies present rather convincing evidence of the benefits of water fluoridation. They illustrate how well-designed preventive medical experiments can be carried out even when randomized control groups are not available.

Example 3: Evaluating the Periodic Multiphasic Health Checkup

An experiment to evaluate the long-term effects of periodic multiphasic health checkups is currently in progress at the Kaiser-Permanente Medical Care Program in northern California. Although the results are only beginning to appear at the time of this writing, this experiment is described to introduce the reader to studies of preventive medical services that go beyond the prevention of single diseases.

It is widely accepted in the United States that annual physical examinations are an important means of maintaining good health. The rationale for annual checkups is that the physician may detect early or asymptomatic disease and initiate treatment before serious consequences develop.

Because of this belief, many persons request and expect annual checkups as part of the medical-care services they receive. Providing checkups to large numbers of patients can consume a substantial proportion of a physician's time—time that might also be used to provide more care of the sick. Because of the growing awareness in this country of the high costs and limitations of physician time and medical care resources, efforts to simplify the checkup are being developed and evaluated. Along these lines, paramedical personnel and automated instruments are being used to assist in examinations in order to save physician time.

Yet the basic question still remains as to just how much overall benefit periodic checkups actually offer. While common sense supports the value of early disease detection and treatment, physicians must also conclude that at least some aspects of checkups (such as listening to the heart and lungs of a young healthy patient every year, year after year) are almost always a waste of time.

The available scientific data on this question are surprisingly limited. A few studies have shown reductions in mortality and in other unfavorable outcomes in groups who received periodic health examinations. However, the comparison groups have not been randomly selected but have been superficially similar populations not receiving examinations. Persons who receive examinations have been shown to be like volunteers and other "cooperators" in that they tend to be more educated, more health-conscious, less prone to

smoke cigarettes, and so on. Thus, serious questions can be raised about the comparability of the examined and nonexamined populations in these earlier studies.

In the Kaiser-Permanente experiment, the control group is quite comparable to the examined, or "study," group. Both groups of over 5,000 subjects were selected on the basis of having certain digits in their medical record numbers, a systematic sampling method that is equivalent to random sampling, since these numbers are assigned in sequence with no relationship to any personal characteristics. These two samples were drawn from a large pool of Kaiser Foundation Health Plan members living in Oakland, Berkeley, and San Francisco, California and aged 35–54 when the study started in 1964. To minimize losses to follow-up, another selection criterion for potential study subjects was that they must have been Health Plan members for at least 2 years, since persons quitting the Plan tend to do so soon after joining.

Each study-group subject has been telephoned and urged to have a multiphasic health checkup every year. Control-group subjects have not been urged or reminded to have these checkups, but, of course, they are entitled to receive this service if they so choose. On the average, 20 to 24 percent of the control group have sought this service each year, and during the first 7 years of the study, the average number of examinations received per subject was 1.34, with 47 percent of control members having received none. In contrast, 60 to 70 percent of the urged study group have been examined annually, and the average number of examinations per subject in 7 years was 3.54, with only 17 percent of study group having had no examinations. Thus the urging has resulted in a considerably larger "dosage" of multiphasic checkups for the study group.

Follow-up of the two groups has consisted of a number of components to measure the development of morbidity, mortality, and disability and to assess the utilization and costs of all medical-care services. Hospitalizations and outpatient visits are tabulated, and the names of all persons lost to follow-up are sent to the state health department for a check against death certificate lists to see if they have died. A questionnaire survey is sent to both groups at approximately 2-year intervals to learn of the development of disability and other pertinent problems.

Whenever possible, assessment of various outcomes is made in

such a way as to avoid bias in favor of study or control group. For example, even though submitting subjects' recent addresses would help the state health department search for deaths, this is not done, since the annual telephone contact with the study group leads to more accurate and up-to-date information about addresses than is available for the control group.

As mentioned, this study is still in progress. Results in the first 7 years show that the checkup program has had an impact on the discovery and diagnosis of a variety of conditions. The older men in the study group, those aged 45–54 when the experiment started, showed some benefit from these examinations in the form of less disability and time lost from work than was experienced by the older control group men. There also appeared to be some reduction in the study group of mortality from conditions that would be expected to be influenced by early detection and therapy, such as hypertension and its complications. Economically, the added costs of the examinations were more than made up for by the greater earning power of the examined group due to their diminished disability and mortality (Cutler et al., 1973, Ramcharan et al., 1973, Dales et al., 1973, Collen et al., 1973).

REFERENCES

Armitage, P., *Sequential Medical Trials*. (Springfield, Ill.: Charles C Thomas, 1960).

Ast, D. B., D. J. Smith, B. Wachs, and K. T. Cantwell. 1956. Newburg-Kingston caries-fluorine study XIV. Combined clinical and roentgenographic dental findings after 10 years of fluoride experience. *J. Am. Dent. Assoc.*, **52**:314–325.

Collen, M. F., L. G. Dales, G. D. Friedman, C. D. Flagle, R. Feldman, and A. B. Siegelaub. 1973. Multiphasic Checkup Evaluation Study: 4. Preliminary cost benefit analysis for middle aged men. *Preventive Medicine*, **2**:236–246.

Cutler, J. L., S. Ramcharan, R. Feldman, A. B. Siegelaub, B. Campbell, G. D. Friedman, L. G. Dales, and M. F. Collen. 1973. Multiphasic Checkup Evaluation Study: 1. Methods and population. *Preventive Medicine*, **2**:197–206.

Dales, L. G., G. D. Friedman, S. Ramcharan, A. B. Siegelaub, B. A. Campbell, R. Feldman, and M. F. Collen. 1973. Multiphasic Checkup Evaluation Study: 3. Outpatient clinic utilization, hospitalization and mortality experience after seven years. *Preventive Medicine*, **2**:221–235.

Dean, H. T. 1956. Fluorine in the control of dental caries. *J. Am. Dent. Assoc.*, **52**:1–8.

Francis, T., Jr., R. F. Korns, R. B. Voight, M. Boisen, F. M. Hemphill, J. A. Napier, and E. Tolchinsky. May 1955. An evaluation of the 1954 poliomyelitis vaccine trials: Summary report. *Am. J. Public Health*, **45**:(No. 5, Part 2)1–63.

Hill, A. B., *Principles of Medical Statistics*, 9th ed. (London: Oxford University Press, 1971), Chap. 20.

Hilleboe, H. E. 1956. History of the Newburgh-Kingston caries-fluorine study. *J. Am. Dent. Assoc.*, **52**:291–295.

Hutchison, G. B., Evaluation of preventive measures, in *Preventive Medicine*, edited by D. W. Clark and B. MacMahon (Boston: Little, Brown, 1967), pp. 39–54.

Ipsen, J., and P. Feigl, *Bancroft's Introduction to Biostatistics*, 2d ed., (New York: Harper and Row, 1970), pp. 180–184.

Ramcharan, S., J. L. Cutler, R. Feldman, A. B. Siegelaub, B. Campbell, G. D. Friedman, L. G. Dales, and M. F. Collen. 1973. Multiphasic Checkup Evaluation Study: 2. Disability and chronic disease after seven years of multiphasic health checkups. *Preventive Medicine*, **2**:207–220.

Schlesinger, E. R., D. E. Overton, H. C. Chase, and K. T. Cantwell. 1956. Newburgh-Kingston caries-fluorine study XIII. Pediatric findings after ten years. *J. Am. Dent. Assoc.*, **52**:296–306.

Smart, J. V., *Elements of Medical Statistics*, 2d ed. (London: Staples Press, 1970) Chaps. 5, 8, 10–12.