

The New England Journal of Medicine

Copyright, 1959, by the Massachusetts Medical Society

Volume 261

SEPTEMBER 17, 1959

Number 12

SOME PROBLEMS ARISING IN PROSPECTIVE AND RETROSPECTIVE STUDIES OF THE ETIOLOGY OF DISEASE*

HAROLD F. DORN, Ph.D.†

BETHESDA, MARYLAND

SIX years ago Bradford Hill, in his Cutter Lecture on the role of observation and experiment in preventive medicine, pointed out that although the experimental method clearly is preferable whenever it can be used, many, if not most, of the problems arising in the investigation of the etiology of disease in human populations, for ethical or other reasons, must be studied by the observation of naturally occurring events. Therefore he did not "repudiate or even . . . underrate the claims of accurate and designed observations." But, he emphasized, in using the method of observation, one should not be content with the examination of readily available observations but "must go seek more facts, paying less attention to technics of handling the data and far more to the development and perfection of methods of obtaining them. In so doing, one must have the experimental approach firmly in mind."¹ These words are as relevant today as they were in 1953.

Why is it that at any given time some fields of investigation seem to be developing rapidly while others appear to be standing still? Two of the most important factors are the available methods of measurement and the general methods of investigation. The history of advances in science could be written in terms of the discovery or development of new technics and methods of measurement. As illustrations, one has only to call to mind the research possibilities opened up by the invention of the electron microscope, the discovery of blood groups and the perfection of methods of tissue culture and serological tests.

The technics, ways of measurement, and methods of investigation used in the epidemiologic study of the etiology of infectious diseases, in contrast to those used in the study of diseases such as cancer, arthritis, cardiovascular diseases, congenital defects and mental diseases, are an apt illustration of this point. It is needless here to enumerate the chemical, bacteriologic, serologic, viral and other technics and methods

of measurement in daily use in the study of the epidemiology of infectious diseases or to describe the greatly increased knowledge of viruses, to mention only one example, that has flowed from the use of these technics and methods during the past two decades.

The contrast with the technics and methods of investigation used in epidemiologic studies of what are usually called the chronic or degenerative diseases is striking. Here, the source of data is frequently the memory of the subject, the recording instrument is a questionnaire or an interviewer, and the unit of data is a word. Moreover, there seems to be a widespread belief that it no longer is necessary to determine as precisely as possible the incidence of a disease among persons with different characteristics. Approximate methods that are less expensive, require less time and hence permit the completion of more studies within a specified period are thought by many to be as satisfactory for this purpose as the more protracted and expensive one of determining the incidence of a disease in a defined population. In view of the important role of technics of measurement and general methods of investigation in the study of the etiology of human diseases, I thought it would be useful to devote this lecture to a discussion of the present status of these in epidemiologic studies of chronic disease, with particular reference to cancer.

It is unnecessary so soon after Bradford Hill's lecture to discuss again the relative merits of the experimental and observational methods of investigation. My remarks will be concerned solely with ways of carrying out observational studies and with some lessons that may be drawn from recent investigations.

THE CASE HISTORY

The most ancient and still the most widely used method of investigating the etiology of human disease is the case history — that is, the study of the characteristics of persons with a disease. The discussion of etiology in medical articles and textbooks is based largely on this method. In addition, it has

*The Cutter Lecture on Preventive Medicine, delivered at the Harvard School of Public Health, Boston, May 19, 1959.

†Chief, Biometrics Branch, Division of Research Services, National Institutes of Health.

been and still is one of the principal ways of investigating the cause of an outbreak of an infectious disease.

There are many achievements to the credit of this method. But almost without exception, these have involved diseases with a specific etiologic agent such as Rocky Mountain spotted fever, Q fever and others caused by micro-organisms. The attempt to reconstruct the causal chain of events preceding the onset of diseases of unknown etiology such as cancer by study of the characteristics and past history of persons with these diseases so far has not been very successful. The most notable discoveries have been the identification of several types of occupational cancer.

One of the earliest illustrations is the description by Percival Pott² of chimney sweep's cancer about 1775. A more recent example is Martland's³ investigation of the subsequent development of cancer and allied diseases among workers in a New Jersey radium-dial-manufacturing plant during World War I. Several other specific types of cancer have been linked by this method to specific occupational exposures.

These illustrations have a common characteristic — they include forms of cancer due to a specific factor encountered by workers in a well defined occupation among whom the number of new cases of disease was sufficiently large to attract attention. In this respect, these forms of cancer resemble outbreaks of infectious disease that are usually investigated by the study of the history of individual cases.

Except for the limited list of malignant neoplasms associated with specific occupational hazards, the study of the characteristics of diseased persons has not contributed much to knowledge of the etiology of cancer. In part, this is due to the long latent period that apparently exists between exposure to a suspected agent and the appearance of clinical symptoms. In addition, the combined effect of multiple factors, no one of which is clearly predominant, may be necessary for the disease to become manifest.

The most serious defect of the case-history method is that it provides no basis for judging the validity of the observations obtained by its use. If the histories of 100 persons with cancer of the liver reveal that 30 have drunk large quantities of alcoholic beverages for many years, should one conclude that alcohol is a precursor of cancer of the liver? Clearly, without a knowledge of the proportion of persons without cancer of the liver who also have drunk large quantities of alcoholic beverages, there is no way of deciding whether the proportion observed among persons with cancer is unusually high or low. In other words, a standard of comparison, usually called a control group, is required.

THE CONTROL GROUP

A standard of comparison is always implicitly, if

not explicitly, required for the evaluation of case-history data used in the study of the etiology of chronic diseases. In Martland's study of radium-dial painters the standard of comparison was his judgment of the frequency with which the diseases observed among the dial painters would appear in usual clinical experience. Ordinarily, a more definite standard of comparison than usual clinical experience is desirable. Recognition of this has led to a modification of the case-history method — namely, the comparison of the characteristics of a group of persons having a specific disease with those of a group of persons not having this disease.

The primary function of this comparison or control group is to provide a basis for evaluating both the assumed explanation of the observed condition and other alternative explanations. For this purpose, the control group should be representative of the population from which the diseased persons came. But this population is usually undefined.

Persons with a disease at a given time are the survivors of an unknown population, some of whose members developed the disease and some of whom did not. Information is lacking about those who did not develop the disease as well as of part of those who did develop the disease.

In most studies the diseased persons included are the first *N* patients admitted to a hospital or group of hospitals after a given date. The characteristics of these persons in facts such as age, sex, race and occupation that might influence the development of the disease studied are not known until after the interview is completed. The population from which the patients came cannot be clearly defined, nor is the process of selection resulting in the choice of the particular patients included well understood. Consequently, the population of which the control group should be representative cannot be defined in advance. The usual procedure is to select the control group from the same source as the diseased persons, in the hope that thereby the two groups will be comparable in all respects except the presence of the disease being studied.

PROSPECTIVE STUDIES

These two methods of investigation — the case-history analysis, and the comparison of diseased persons with a control group — are retrospective in the sense that the subjects are selected after the disease has developed. An attempt is then made to define the population from which the subjects have come and to discover the chain of events antecedent to the development of the disease. The difficulty of doing this, the lack of knowledge concerning possible selective processes that might bias the observations and the consequent uncertainty concerning the validity of the conclusions gave rise to the use of a third method of investigation, prospective studies.

This method starts with a defined population that is kept under observation to determine the risk that a particular disease will develop in persons with specified characteristics. The characteristics that are thought to influence the occurrence of the disease can be defined and described before the period of observation begins. This ensures that the same information will be available for those in whom the disease develops as for those in whom it does not. A known population at risk is available so that measures of the probability of the occurrence of the disease can be computed.

Such studies need be prospective only in the sense that it is possible to define and identify the exposed population before observations are made. If the necessary records are available, a population could be defined as of some past date, and a study made of its experience with respect to the occurrence of a disease after that date. Investigations of the mortality of insured persons with various medical impairments have been done in this way.⁴

COMPARISON OF PROSPECTIVE AND RETROSPECTIVE METHODS

These attributes of the prospective method of study have given rise to a widespread belief that it is superior to either of the retrospective methods in the sense that greater confidence can be placed in the results. Recently, however, several investigators have expressed the opinion that the retrospective method, in particular the comparison of diseased persons with a control group, can yield data of equal validity to those obtained by prospective studies. In addition, retrospective studies are less expensive, can be completed more quickly and are more easily done so that presumably more scientific knowledge can be obtained per dollar per day. This opinion is apparently shared by a majority of investigators since the number of retrospective studies of the etiology of chronic disease reported in recent years in the United States far exceeds the number of prospective studies.

The primary basis for this opinion seems to be a comparison of the results of retrospective and prospective studies of the relation of smoking to cancer of the lung. During the past two decades the results of three large-scale prospective studies and more than twenty retrospective studies have been reported. All the retrospective studies have shown a larger percentage of heavy smokers and a smaller percentage of nonsmokers among male patients with cancer of the lung than among control patients. Each of the prospective studies has shown that the death rate from cancer of the lung is higher among cigarette smokers than among nonsmokers and that the relative difference in the death rates increases with an increase in the average daily number of cigarettes smoked.

The following conclusion based on a comparison of

the results of a retrospective study of smoking and lung cancer conducted at Roswell Park Hospital with those obtained in the prospective study carried out by Hammond and Horn of the American Cancer Society typifies the opinion apparently shared by many investigators:

It is also evident that, apart from the closeness of the actual rates in the two studies, both methods of study have led to the same conclusions with respect to the relative incidence of lung cancer among nonsmokers, other than cigarette smokers, and cigarette smokers. . . . The similarity of results obtained by the whole-population study of Hammond and Horn, in which smoking histories were obtained from more than 200,000 persons, with the hospital samples, in which 2855 patients were studied, is of added interest because of the obvious fact that epidemiological studies of hospital populations can be done much more readily, more frequently, and with more careful attention to a large number of variables, than can whole-population prospective studies. . . . Provided each is well designed, there would appear to be no important theoretical reasons for preferring one type of study to the other.⁵

It is important to know whether this statement is generally true. If it is, the belief, held by many, that prospective studies give a more reliable measure of the risk of the development of a disease is not valid, and the simpler, less expensive and quicker retrospective method of study should be the preferred method of investigation. I am not convinced that the statement is generally true, and I shall illustrate the reasons for my disbelief by some data obtained in a study of cancer of the cervix in Jewish and non-Jewish females conducted by the National Cancer Institute.

Clinical literature contains much speculation concerning the etiology of cancer of the uterine cervix derived from a review of the history of patients. Several investigators have checked these suggestions by retrospective studies of the characteristics in cases of cervical cancer and a corresponding control group. Three of the most comprehensive studies are those by Lombard and Potter,⁶ Wynder and his associates⁷ and Jones, Macdonald and Breslow.⁸

Lombard and Potter⁶ interviewed patients with cancer of the cervix in two Massachusetts State Cancer Hospitals and six different groups of persons without cancer of the cervix, including hospital patients, friends of the patients with cancer and persons selected by a household community survey. After reviewing the data collected concerning 29 factors, the authors concluded, "The strong correlations that exist between cancer of the cervix and marriage before the age of 20, individuals divorced or separated at any time, unrepaired lacerations, last child born to woman before age 25, and syphilis indicate that these variables are of etiological significance."⁶

Wynder et al.⁷ interviewed Jewish, non-Jewish and Negro females in hospitals in New York City and St. Louis. They reported a strong positive association of the following factors with cancer of the cervix: early age at first marriage; early age at first coitus;

multiple marriages; and lack of circumcision of sexual partners.

Jones and his co-workers⁶ studied patients in the Los Angeles County Hospital and from private practice. Control cases were selected from the same sources. No significant difference between patients and controls was found in dietary deficiency, estrogen excretion levels, menstrual patterns, hygienic practices, use of contraceptives and frequency and dura-

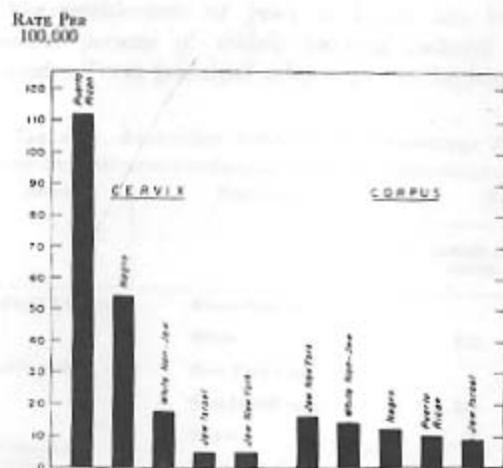


FIGURE 1. Incidence of Cancer of the Uterine Cervix and Uterine Corpus in Selected Population Groups — Number of New Cases per 100,000 Jewish Females in Israel, 1952-53, and per 100,000 Puerto Rican, Negro, Jewish and Non-Jewish White Females Respectively in New York City, 1952 (Rates Are Age Adjusted).

tion of coitus. Contrary to Wynder, Jones did not find that a higher proportion of women with cancer of the cervix than of the control cases reported un-circumcised sexual partners. Lombard apparently did not include circumcision status among the factors investigated in his study. Early age at first marriage was found to be positively associated with the development of cervical cancer in each of the three studies.

Many persons have reported that cancer of the cervix occurs much less frequently among Jewish than among non-Jewish females. Although this finding is based almost entirely on a review of hospital case records and autopsy statistics, it has been reported from several countries and has been widely accepted as true.

A few years ago the National Cancer Institute began a series of studies to determine the validity of the reported association between various factors and cancer of the cervix. Each study had three aspects: the determination of the incidence of cancer of the cervix in defined populations; the collection of case histories of patients with cancer of the cervix and of a control group of persons without cancer of the cervix; and the verification of the diagnosis of each case of cancer by an independent review of the available pathologic tissue.

The objective of the first study was the collection of these data for Jewish and non-Jewish populations since available information suggested the largest difference in the incidence of cervical cancer might be found in these two groups. Subsequent studies included the Negro population of New York City and Washington, D.C.

In each study an effort was made to obtain a report on every case of cervical cancer first diagnosed during the study period in the population studied, to compute the incidence rate and to determine the proportion of new cases that were interviewed. An attempt was made to interview every new patient with cancer in the Jewish population of Israel and New York City. The number of new cases of cancer of the uterus in the non-Jewish white population of New York City was considerably larger than that in the Jewish population; consequently, only a sample of these patients were interviewed. In both areas the study included cases of cancer of the uterine corpus and, in Israel, cases of cancer of the ovary and breast

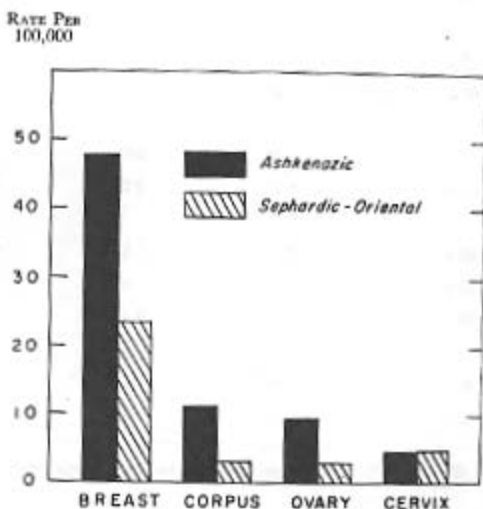


FIGURE 2. Incidence of Breast and Genital Cancer among Ashkenazic and Sephardic-Oriental Jews in Israel, 1952-53 — Number of New Cases per 100,000 Females, Adjusted for Age.

This method of investigation — namely, the collection of data concerning every newly diagnosed case of cancer in a defined population — permitted a comparison of the conclusions drawn from an analysis of factors associated with cancer of the cervix based on a retrospective study with incidence rates for the same population.

The incidence rates shown in Figure 1 reveal a wide range in the incidence of cancer of the cervix in the various population groups identified, a small variation in the incidence rate of cancer of the corpus, a lack of correlation between variation in the incidence of cancer of the cervix and cancer of the corpus (a fact suggesting that different etiologic agents exist for each type of cancer) and very low and essentially

identical incidence rates for cervical cancer among Jewish females in Israel and New York City. The incidence of cervical cancer was twenty times greater among Puerto Ricans than among Jews, with the rates for Negro and non-Jewish white females lying between these two extremes.

The resettlement of Jews in Israel has brought together persons of widely varying cultural backgrounds. Three principal subgroups can be identified

the same population during the same period, both the retrospective and the prospective methods of studying the association between various factors and a disease should theoretically lead to the same conclusions.

The studies conducted by Lombard and Potter,⁶ Wynder et al.,⁷ Jones and his associates⁸ and the National Cancer Institute investigated some of the same factors thought to influence the development

TABLE 1. Association between the Percentage Married Before Twenty Years of Age and Cancer of the Cervix.

STUDY	POPULATION	NO. OF CASES		PERCENTAGE MARRIED BEFORE AGE 20		RELATIVE RISK OF CANCER OF CERVIX*
		CANCER OF CERVIX	CONTROL	CANCER OF CERVIX	CONTROL	
Lombard & Potter ⁶	Massachusetts					
	White	523	298-565	44.6	16.3-23.7	4.13-2.59
Wynder et al. ⁷	New York City					
	Non-Jewish white	129	302	51	31	2.32
	Negro	56	95	71	51	2.35
National Cancer Institute	New York City					
	Non-Jewish white	320	569	40.0	36.7	1.15
	Jewish	45	341	22.2	23.4	0.93
	Israel					
	Total Jewish	48	1195	52.1	22.8	3.68
	Ashkenazic	32	966	34.4	17.0	2.56
	Sephardic-Oriental	16	229	87.6	47.6	7.78
Jones et al. ⁸	Los Angeles					
	Ward	153	153†	66	56	1.53
	Private	149	149†	42	35	1.35

*Ratio of estimated incidence rate for cancer of cervix among females married before age 20 to rate for females married after age 20.
†Equivalent number after age standardization.

—the Ashkenazic, literally German Jews but rather loosely referring to those coming from Northern and Central Europe, the Sephardic or Spanish and Portuguese Jews, coming principally from the countries bordering the Mediterranean, and the Oriental Jews, coming from Asia Minor. The number of cases of cancer among Sephardic and Oriental Jews during the two-year period of study was not large enough to warrant the computation of incidence rates for each group separately. Rates for both groups combined and for Ashkenazic Jews are shown in Figure 2.

The most remarkable feature of these data is the similarity in the incidence rates for cancer of the cervix in the two groups, in contrast to the disparity in the incidence rates for cancer of the breast, corpus and ovary. No significant variation was found in the incidence of cancer of the cervix in any of the separate groups of Jews identified in New York City and Israel.

If both the cancer and control cases are respectively representative samples of the newly diagnosed cases of cancer in a defined population during a specified period and of the persons without cancer in

of cancer of the cervix. A higher percentage of cervical cancer than of control patients married earlier than twenty years of age was found in each study. Data showing the association of the percentage married before age twenty and the development of cervical cancer are presented in Table 1.

In each of the nine population groups, except Jewish females in New York City, as reported by the National Cancer Institute, a larger percentage of women with cancer of the cervix than of women without cancer of the cervix were first married under twenty years of age. However, it is virtually impossible from an inspection of these percentages to determine the closeness of the association between the percentage married before age twenty and the development of cervical cancer or whether this association is stronger in one group than in another. For this purpose a measure of the risk of the development of cancer of the cervix among females married before age twenty and among females married after age twenty is necessary.

Such a measure can be estimated from the data in Table 1 if it can be assumed that the cancer and con-

trol cases are representative of the corresponding groups in the population from which each is a sample. In the absence of a knowledge of the incidence rate in the population studied, the only measure that can be computed is the relative risk that cancer of the cervix will develop.⁹ This is shown in the last column on the right in Table 1.

The number, 3.68, for the Jewish population of Israel means that the risk that cancer of the cervix will develop in females who marry before age twenty is 3.68 times the corresponding risk for females who marry after age twenty. The variation in the estimated risk of cancer of the cervix among the several population groups—a variation not directly related to the size of the percentage marrying before age twenty—emphasizes the difficulty in the interpretation of retrospective studies pointed out above (namely, that the persons included are usually selected by an unknown method of sampling from an unspecified population).

Lombard and Potter⁶ used six different control groups. The data for each show that women who marry before age twenty have a greater risk of cervical cancer than women who marry after this age, but the estimate of the relative risk varies from 2.6 to 4.1. The highest estimate, 7.78, is for the Sephardic-Oriental Jews in Israel and is three times that for Ashkenazic Jews. The lowest estimates were found by the National Cancer Institute for Jewish and non-Jewish white females in New York City and by Jones et al.⁸ in Los Angeles.

As mentioned above, in the study by the National Cancer Institute, an effort was made to interview every patient with a new cancer diagnosed during the study period in the Jewish populations of Israel and New York City. However, some patients died before arrangements could be made for an interview, others were too sick to be interviewed, a few refused to co-operate, and in Israel some could not be interviewed owing to language difficulties. As a result, 23 per cent of the new patients with cancer of the cervix in the Jewish population of New York City and 83 per cent of those in Israel were interviewed. No information is available concerning the proportion interviewed in the studies by Lombard and Potter,⁶ Wynder et al.⁷ and Jones, Macdonald and Breslow.⁸

The six control groups used by Lombard and Potter were selected from hospital patients and by a community household survey. In the other three studies the controls were selected from the same source as the patients with cancer. In all studies they were a very small percentage of the total number of females without cancer.

Much more serious than the variation in the estimates of relative risk obtained from the nine groups is the fact that these estimates are not related consistently to incidence rates. Incidence rates are not

available for females classified by age at first marriage. However if there is a strong positive association between early age at first marriage and the development of cervical cancer, as the data in Table 1 suggest, the incidence rate for cancer of the cervix should be higher for populations in which a high percentage of females marry before twenty years of age. But this obviously is not consistently true.

The percentage of females marrying before age twenty is more than twice as great for Sephardic-Oriental Jews as for Ashkenazic Jews for both cancer and control cases; yet the incidence of cervical cancer does not differ. Neither do the incidence rates differ for New York Jews and for Israeli Jews although the percentages marrying before age twenty differ widely. The incidence rate for cancer of the cervix among Negroes in New York City is higher than that for any group except Puerto Ricans. The percentage marrying before age twenty is also higher than that for each group except the Sephardic-Oriental Jews in Israel, but the latter has a lower incidence rate.

The lack of agreement between the measure of relative risk of cervical cancer obtained from the retrospective interviews and incidence rates as well as the variation in the estimates of relative risk among the retrospective studies is due largely to the fact that the cancer and control cases were not representative samples of the unknown population from which they came. But how can it be determined in any specific group that the cases studied are not representative so long as the population from which the cases originated as well as the selective process that resulted in their choice is unknown?

The inconsistency is probably due in part to the fact that the percentage marrying before age twenty is not a primary etiologic factor for cervical cancer. There is no reason to believe that the performance of a marriage ceremony in itself has any direct effect upon the risk that cancer of the cervix will develop. Age at marriage varies widely among different population groups and also may vary over time within a given group. Early age at marriage for females has been a prominent characteristic of the culture of the Sephardic-Oriental Jews now living in Israel whereas the cultural background of the Ashkenazic Jews has favored a later age at first marriage.

During the past two decades the percentage of females marrying before the age of twenty in the United States has increased sharply. In 1940 the Bureau of the Census reported that 22 per cent of females eighteen or nineteen years old were married; in March 1958 this had increased to 34 per cent.¹⁰ If females who marry before twenty years of age are more likely to develop cancer of the cervix, the incidence rate soon should begin to increase since women who married before that age during the early 1940's are now between thirty and forty years of age, the age

period during which incidence rates first become numerically large.

The one trait, marriage before age twenty, found by each of the four studies to be positively associated with cervical cancer is not consistently related to variation in the incidence rate. The studies yielded contradictory results in several other respects.

Wynder et al.⁷ concluded that lack of circumcision of sexual partners appreciably increases the risk that cancer of the cervix will develop. The studies by Jones and his co-workers⁸ and the National Cancer Institute reported no association between lack of circumcision and cancer of the cervix. Jones, Macdonald and Breslow found no association between the number of sexual partners and cervical cancer. Wynder et al. and the National Cancer Institute found an association, but the relative increase in risk for the non-Jewish population of New York City estimated from the data of the latter study was only 16 per cent. A higher proportion of patients with cervical cancer than of control patients were divorced or separated in the studies of Lombard and Potter, Jones and his associates and the National Cancer Institute. However, the death rate of divorced or separated women is higher than that of married women both for all causes combined and for a number of specific diseases so that this finding probably reflects only the general health status of such women.

It is difficult to see how age at marriage per se, as well as other factors such as the fact of divorce or separation and low economic status that have been identified in retrospective studies, can directly influence the development of cancer of the cervix. Early age at marriage may be related to some presently unknown factor that does directly influence the development of cancer of the cervix. If so, an effort to identify this unknown factor would be preferable to a repetition of studies showing that a higher proportion of women with cervical cancer than of women without cervical cancer are married before twenty years of age.

DISCUSSION

I do not wish to give the impression that I reject retrospective studies as a method of investigating the etiology of chronic diseases. The retrospective method, in theory, can provide data of reliability and validity comparable to that obtained from prospective studies. But, as usually applied, it does not do this. The fundamental defect of many retrospective studies is that they are based on an unspecified sample of persons chosen by an unknown method of sampling from an unidentified population.

The fact that retrospective studies conducted in this manner sometimes lead to the same conclusions as more expensive and prolonged prospective studies is not a sound basis for assuming that this generally

will be true. A purposive or haphazardly chosen sample may sometimes lead to the same conclusions as a representative sample selected by probability methods, but experience has demonstrated that this is generally not true and that there is no way of identifying in advance the occasional instance when it will be true.

The retrospective method can be particularly useful in the study of diseases of a very low incidence and in the tentative exploration of hypotheses when not much is known about etiology of a disease. But if it is to be useful it must provide a valid basis for generalizing from the observations obtained.

The primary purpose of epidemiologic studies of the etiology of disease is to discover generalizations that will hold true for a population larger than that on which the study is based. This is equally true for retrospective and prospective studies. In this sense all studies are based on samples.

If valid inferences are to be drawn from a sample, two requirements must be fulfilled: the sample must be representative of the population to which inferences are to be drawn; and the data collected must be of sufficient reliability and validity to answer the hypotheses to be tested. It is essential to fulfill both requirements. Neither requirement should be relaxed without a careful weighing of the consequences, since confidence in the results of the study will thereby be weakened. Too frequently there is a temptation to relax both requirements because it is cheaper, more convenient or quicker to use readily available data even though these may be of doubtful validity and cannot be referred to any known population.

There is no way of ensuring that a single specific sample selected by any method will be representative of a given population in the sense that the same conclusions would be drawn from a study of the sample as from a study of the entire population from which it is taken. Nevertheless, experience has demonstrated that some methods are more likely than others to accomplish this objective.

The first essential is that the sample of persons chosen for study should be selected by a known method from a defined population. This is the requirement most commonly violated by retrospective studies. It is true that every group of persons may be thought of as a sample of some population. A study of the characteristics of the sample may result in a general conception of the population from which it could have come. But this deduction is of uncertain validity and usually cannot be demonstrated objectively. Hence, the results of studies based on samples from undefined populations cannot be regarded with the same confidence as the results of studies based on samples drawn by probability methods from a defined population. Before any study is undertaken, careful consideration should be

given to the question of whether the objectives can be fulfilled by observations based on the sample of persons selected.

The only method of sampling that is known to produce a representative sample in most studies is that of probability sampling. This method has two principal advantages: it permits the computation of a valid error of sampling; and, more important, it eliminates the bias in the selection of the persons to be studied that may occur from purposive selection or from accepting the first *N* patients that are admitted to a single hospital.

An alternative to selecting a sample of cases for interview is to include all new cases developing in a specified population during a defined period. This leaves unresolved the problem of whether the population chosen has been subject to some selective factors in the past that may bias or limit the generality of any conclusions, but it does eliminate the uncertainty about the method of selection of the cases with disease although not necessarily about the selection of the control cases.

A number of steps can be taken to improve the methodology of retrospective studies so that they will provide a more valid basis for generalization:

A defined population should be chosen for study. Otherwise, there is no way of determining how the selective factors that have produced the population may affect the validity of the observations to be made. The most easily defined populations are persons living in a delimited geographic area such as a city, village or township, members of a specific occupation or profession such as bus drivers or physicians, employees of a company, members of a labor union and so forth.

All newly diagnosed cases of disease or deaths occurring in the defined population during a specified period should be included.

Whenever possible, one should use objective measures of the disease studied and of the factors or traits thought to be associated with the disease. The verification by physical examination of statements concerning circumcision has thrown doubt upon the validity of conclusions concerning the relation of lack of circumcision to the development of cancer of the cervix based on data obtained by interviewing wives.¹¹ It has also shown that the dichotomous classification of males as circumcised or not circumcised is an oversimplification of the condition as it, in fact, exists.¹²

More than one type of control group should be used. Unless the same association between the factor or trait in question and the existence of the disease studied is found for each control group, the validity of the association is questionable.

The study should be repeated in several different populations in which the supposed etiologic factors vary widely in intensity or degree. The investiga-

tion of cancer of the cervix in populations with a wide variation in age at marriage is an illustration of this point. Until an observed association is verified by replication, it is merely a single observation and should not be accepted as established.

Primary etiologic factors should be sought. A review of published retrospective studies of cancer shows that the usual pattern is to make one study of a particular type, — for example, cancer of the cervix, — publish a paper and then study another type of cancer. Most of the suggested associations — early age at marriage, divorce or separation and low economic status — clearly are not primary etiologic factors. It is very unlikely that a single study of a disease will uncover primary etiologic factors. These are found by repeated investigations of the same disease, building upon and extending existing knowledge.

One should search for populations known to be exposed to a condition suspected of causing a disease. Attempts to discover possible occupational or industrial hazards that influence the development of specific forms of cancer by a retrospective study of patients with the disease from the general population have shown that this method of study is too diffuse to be likely to reveal specific carcinogenic hazards.¹³ Most knowledge of occupational hazards that cause cancer has been obtained by observation of workers in a specific occupation or performing a clearly defined operation. Usually, the number of workers exposed to a hazard is not large so that retrospective studies of workers with a history of having worked in many rather vaguely defined occupations and having only a sketchy knowledge of possible exposures are unlikely to yield definitive results. A prospective study of workers who are performing a specific job and for whom exposure to the suspected hazard can be verified is clearly the preferred method.

One should verify the strength of an observed association between an etiologic factor and a disease by a prospective study yielding either incidence rates or mortality rates for the disease. Retrospective studies do not give a direct measure of the risk that a disease will develop in persons with or without a specified trait. A measure of relative risk can be computed subject to certain assumptions concerning the representativeness of the persons included in the study, but an analysis of computations from a large number of retrospective studies shows that little confidence can be placed in the estimate of relative risk obtained from the data of a single retrospective study. It is true that examples, such as that cited by Levin et al.,⁵ can be found for which the relative risk of a disease computed from a retrospective study agrees closely with that computed from a prospective study, but there is no way of being sure of this without a

prospective study. There is no substitute for measuring as precisely as possible the risk of a disease among members of a defined population.

Many believe that only trained investigators can plan and conduct controlled experiments whereas even the man on the street can record and interpret observations of naturally occurring events. But, in fact, it is equally difficult, if not more so, to make accurate, reliable observations of naturally occurring events than it is to carry out planned experiments. To unravel the tangled skein of causation of disease by the observation of man in his environment requires the utmost persistence, a profound skepticism of the obvious, an alertness for the selective factors that have produced the most readily available subjects for study, the ability to penetrate beneath the surface of the observations, a strong distrust of what is said or seen since the unchecked human memory is probably the most unreliable research instrument in existence and a willingness to spend long hours upon the development and perfection of methods of making observations coupled with the patience to await the occurrence of the succession of events most relevant to the hypotheses to be tested. The more inefficient the method of investigation, the greater the necessity for a critical attitude toward the selective factors that may have produced the subjects studied and toward the reliability and validity of the observations.

How else but by observation upon man himself can one hope to find clues to the etiology of his diseases? But these observations should be accurate and

well designed, obtained by a persistent search rather than by a quick walk through a hospital ward as if one were picking up flotsam by casually strolling along the beach. Nature guards her secrets well and throws up deceptive camouflages. The method of observation can be very useful in penetrating this camouflage and in finding the concealed secrets but only if it is well directed and controlled.

REFERENCES

- Hill, A. B. Observation and experiment. *New Eng. J. Med.* 248:995-1001, 1953.
- Pott, P. *The Chirurgical Works of . . .* 2 vols. Vol. 2. Dublin: J. Williams, 1778. P. 403.
- Martland, H. S. Occurrence of malignancy in radioactive persons: general review of data gathered in study of radium dial painters, with special reference to occurrence of osteogenic sarcoma and interrelationship of certain blood diseases. *Am. J. Cancer* 15:2435-2516, 1931.
- Society of Actuaries. *Impairment Study, 1951*. 300 pp. Chicago: The Society, 1954.
- Levin, M. L., Kraus, A. S., Goldberg, I. D., and Gerhardt, P. R. Problems in study of occupation and smoking in relation to lung cancer. *Cancer* 8:932-936, 1955.
- Lombard, H. L., and Potter, E. A. Epidemiological aspects of cancer of cervix. II. Hereditary and environmental factors. *Cancer* 3:960-968, 1950.
- Wynder, E. L., Cornfield, J., Shroff, P. D., and Doraiswami, K. R. Study of environmental factors in carcinoma of cervix. *Am. J. Obst. & Gynec.* 68:1016-1052, 1954.
- Jones, E. G., Macdonald, I., and Breslow, L. Study of epidemiologic factors in carcinoma of uterine cervix. *Am. J. Obst. & Gynec.* 76:1-10, 1958.
- Cornfield, J. Method of estimating comparative rates from clinical data: applications to cancer of lung, breast, and cervix. *J. Nat. Cancer Inst.* 11:1269-1275, 1951.
- United States Census Bureau, Commerce Department. *Marital Status and Family Status, March 1958*. 15 pp. Washington, D.C.: Government Printing Office, 1958. (Current Population Reports, Series 20, No. 87.)
- Lilienfeld, A. M., and Graham, S. Validity of determining circumcision status by questionnaire as related to epidemiological studies of cancer of cervix. *J. Nat. Cancer Inst.* 21:713-720, 1958.
- Dunn, J. E., Jr., and Buell, P. Association of cervical cancer with circumcision of sexual partner. *J. Nat. Cancer Inst.* 22:749-764, 1959.
- Kraus, A. S., Levin, M. L., and Gerhardt, P. R. Study of occupational associations with gastric cancer. *Am. J. Pub. Health* 47:961-970, 1957.

LATEX-FIXATION TEST IN RHEUMATOID ARTHRITIS*

II. Characterization of the Thermolabile Inhibitor by a Serologic Study

ADALBERT F. SCHUBART, M.D.†

BOSTON

THE latex-fixation test of Singer and Plotz¹ was designed as a diagnostic tool for the detection of rheumatoid arthritis. The positive reaction in this test depends upon the interaction of human gamma globulin with, presumably, various serum components, among which the rheumatoid factor is outstanding.

*From the Department of Medicine, Harvard Medical School, the Medical Services, Massachusetts General Hospital, Boston, and the Department of Medicine, University of Maryland School of Medicine, Baltimore.

This is Publication No. 254 from the Robert W. Lovett Memorial for the Study of Crippling Disease. Grants in support of these investigations have been received from the Commonwealth Fund, New York City, the National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, United States Public Health Service (Grants No. A-1064 and A-2123) and Eli Lilly Company.

†Assistant professor of medicine, University of Maryland School of Medicine (on leave of absence); (part of this work during the tenure of a traineeship of the National Institute of Arthritis and Metabolic Diseases).

The polystyrene latex particles act only as carriers for the gamma globulin and as indicators for the precipitin reaction that occurs on the surface of the biologically inert particle.

The interaction of the rheumatoid factor and gamma globulin may be inhibited naturally in a given serum to the extent that aggregation of the particles no longer occurs. This inhibition was mainly present in low serum dilutions resulting in a so-called "prozone inhibition."²⁻⁶ Previous observations^{7,8} have indicated that this natural inhibitor is relatively thermolabile and that it possesses some of the properties of serum complement.

It is the purpose of this paper to present the results of serologic studies performed to characterize further