EPIDEMIOLOGY IN THE UNITED STATES AFTER WORLD WAR II: THE EVOLUTION OF TECHNIQUE

MERVYN SUSSER

PROLOGUE

Epidemiology as a discipline is made unique by the combination of a way of looking at things—the theory and methods of population studies—and the subject of its studies, states of health. In my view, what distinguishes the present era from the past are the concepts which guide epidemiologists. It is not so much the particulars of their discoveries about states of health as a change in theories of disease and the consequent creation of a body of methods to test those theories. These theoretic and technical ideas evolve from changes in society and from concurrent changes in the configuration of health and disease.

In this paper, I focus on the theories and techniques used to cope with the changed spectrum of diseases. My primary source is precisely the literature of epidemiology on theory and method. In the period under review, theory and method have become predominant in the education of epidemiologists. It is this literature, and the growing numbers of women and men who teach and learn it, which has given epidemiology standing as a mature discipline.

There are other signs of a coming of age. Although epidemiology has long been assigned a place as a basic science of public health and preventive medicine, in the past its practitioners have been products and adjuncts of medicine. Virtually the only route of entry into epidemiology was from medicine and through public health. Rarely did a substantively oriented nonmedical scientist—a social scientist like Edgar Sydenstricker in 1915, or a few years later in England, a statistician like Austin Bradford Hill—enter the field.

A measure of the majority of the discipline is that since the early 1970s epidemiologists have been able to pursue doctoral studies without having acquired a primary degree beyond the baccalaureate in some other field. Academic programs produce trained professionals capable of pursuing research autonomously. Such an epidemiologist is competent in statistics but not a statistician; has a grasp of concrete biomedical reality without being a clinician; and comprehends society and social structure without being a sociologist or anthropologist. Ideally, the new-minted epidemiologist should be capable of doing better than any one of these in pursuing the epidemiologic task.

This new breed had to await the synthesis of a solid disciplinary foundation. Only recently have epidemiologic writings assumed the definable structure of a coherent whole and provided the needed foundation. Thus, the coming of age of epidemiology as a discipline signifies a body of research and literature that is sufficient for academic studies in depth. Methods adequate to the various tasks of epidemiology have been...
devised and clarified, and the choice among them depends not only on intuition, natural ability, and a knowledge of medicine, but also on specialized training in a particular field. In consequence, the autodidact is a dying species. Great epidemiologists of the past, the creators of the discipline, had little special training. The mark of their greatness is precisely that they were the innovators, self-taught of necessity.

Professional organizations and their memberships testify to these advances in maturity. Both the Epidemiology Section of the American Public Health Association (long the main professional organization for American epidemiologists) and the American Epidemiological Society (a closed and self-elected group of senior epidemiologists) are more than 50 years old, having been founded in 1927 and 1928, respectively. The Society for Epidemiologic Research, however, founded by Abraham Lilienfeld, Brian MacMahon, and Milton Terris in 1968, was the first open organization devoted specifically to epidemiology. By 1982, its membership had increased from 378 to 1,761, including about 400 students. (Most recently, an American College of Epidemiology has been formed. Epidemiology is an applied as well as an academic discipline. Consonant with the history of other practicing professions, the College can be seen as an attempt to establish a professional identity, common standards, and professional control over practice and qualifications.)

Epidemiology today is fully institutionalized through academic departments in major universities, an assembly of common texts, doctoral programs capable of producing faculty with primary training in the discipline, and a research output that has a significant and continual impact on research and practice in the health field. This is a phenomenon of the post-war era. One may ask why so late? and why then?

Delayed maturation

To recognize the existence of maturational delay, one might compare the lapse of time, on the one hand between the work of William Harvey in the third decade of the seventeenth century and the founding of physiology as an academic discipline in France by the mid-nineteenth century, and on the other hand between the work of John Graunt and William Petty in the mid-seventeenth century and the like founding of epidemiology as a discipline. In this comparison of the basic sciences of medicine and of public health, epidemiology lagged by almost a century.

In fact, the study of associations between environmental factors and disease had gained apace early in the 19th century. Workers like Louis René Villerme in France, and William Farr, John Snow, John Simon, and William Guy in England not only enumerated and quantified their data but also studied communities and populations especially selected to illuminate such areas as the effects of drainage, housing, occupation, and nutrition.

Especially in England, industrialization, urbanization, and growth together created the need to confront a changing pattern of disease that flourished in the rapidly growing urban slums. The challenge excited the medical scientists of the time. John Simon, as Chief Medical Officer of the General Board of Health from 1855 to 1858, could draw the most brilliant young doctors in the country into his epidemiologic team (1). By the last quarter of the century, however, the pace had slowed. An unfavorable political climate frustrated the efforts of public health workers (2, 3). At the same time, another brilliant—but differently focused era opened for public health with the discoveries of microbiology.

I am not alone in thinking that some of the maturation lag of epidemiology can be attributed to the enormous successes of the germ theory (4–7). In terms of the insight of Kuhn (which is a guiding idea in this paper), this was the acceptance by the scientific community of a new paradigm (8), one that eliminated its chief rival, the miasma theory (6). As a result, the Henle-
Koch postulates—laboratory-based criteria for judging the causal relationship of organism to lesion—displaced the population-based inferences typical of the work of the earliest modern epidemiologists. In the 1880s, Koch began to apply the discoveries of Pasteur to the elucidation of specific human infections. Quite soon, the focus of epidemiology was reduced to the pursuit of specific agents, singular causes, and the means of preventing their consequences.

The population concepts and quantitative skills from the first half of the nineteenth century were scarcely maintained in the first half of the twentieth century. In England, some rare souls like J. Brownlee, M. Greenwood, W. O. Kermack, and A. G. McKendrick continued the tradition (6). In the United States, these notions of epidemiology were sustained by a few like Charles V. Chapin and Wade Hampton Frost, who in 1921 occupied the founding chair of the discipline at the Johns Hopkins School of Hygiene and Public Health. In distilling general principles of the epidemiology of specific infections, Frost expressed concisely the interactions of microorganisms, populations, and environmental conditions (9). In studying particularly chronic infectious disease like tuberculosis, and infections by unknown organisms like poliomyelitis, he was responsible for theoretic formulations that have been adapted and developed in the current phase of the study of chronic noninfectious disease.

Parasitologists like Ronald Ross in Britain and Theobald Smith in the United States also helped somewhat to broaden the focus of attention. Ross' Theory of Happenings of 1910 aimed at the prevention of malaria was the beginning of multivariate modeling in epidemiology (6). The life cycles of parasites that caused human disease involved reciprocal relationships between agent, host, and environment that could not be ignored. Smith elaborated the concept of host and parasite in his book Parasitism and Disease in 1934. Others used these ideas of reciprocal interactions between multiple factors to launch a new phase in epidemiology that challenged the dominance of the germ theory.

This new phase restored the relevance of population research based on quantitative methods. It opened in Britain with population studies of chronic disease in academic departments of social medicine. J. A. Ryle at Oxford in 1942, and soon after, F. A. E. Crew at Edinburgh, A. C. Stephenson at Belfast, T. McKeown at Birmingham, and others were appointed to founding chairs in social medicine. J. N. Morris headed the Medical Research Council Social Medicine Research Unit, and Austin Bradford Hill at the London School of Hygiene was active in teaching and research.

Take-off in the United States came later (10). Although the US Army Epidemiological Board did crucial work during World War II in such areas as the control of malaria, the study of hepatitis, and yellow fever and influenza vaccines, the very strength of this work and its focus on infection may have retarded the shift to a focus on chronic disease.

**Transition**

For the new phase, World War II is a convenient watershed. Leading epidemiologists perceived and prepared themselves to meet the *epidemic transition* (11) from the predominance of infectious disease to that of chronic disease.

This transition was the outcome of objective and continuing change in society at large. The Great Depression did not for long interrupt a steady increment in economic productivity that led to higher living standards, better nutrition, and large demographic changes. The population distribution shifted substantially toward older ages. The decline in infant and childhood mortality increased the numbers surviving to middle and old age; the decline in births reduced the numbers of the young. With these age changes came the now familiar increment in age-related disease such as heart disease, stroke, cancer, and senile...
dementia. Alongside the economic and demographic transformation, the scientific application of sanitation and vaccines and of chemotherapy and antibiotics undoubtedly contributed to the conquest of the more severe effects of infectious diseases.

While the infectious diseases declined among the forces of mortality, new and mysterious diseases, so-called "diseases of civilization," began to emerge. Middle-aged males in particular began to suffer and die from them at an unprecedented rate—after the turn of the century, from peptic ulcer, and in the 1920s and 1930s, from coronary heart disease and lung cancer (12). Hypertension and mental disorders were also thought to be on the rise, although no good population data were available to bear this out.

At the time of this epidemic transition, in medicine the prevailing explanatory paradigm for chronic disorder was that of degenerative disease. These diseases were seen in the main as intrinsic failures of the aging organism, brought to attention by increased survival and the rise in the aged population. Distinct from the orthodox medical view, however, a novel perception among epidemiologists entrained a line of research which was to establish a different paradigm for chronic disease. The view that chronic diseases had environmental causes that were potentially preventable reflected the optimism and enthusiasm for reform that prevailed at the end of World War II. After a long and uphill struggle, that paradigm is widely accepted, in medicine as well as in public health. At the same time, and I believe as a result, epidemiology has achieved the respectability of a needed discipline in the biomedical armamentarium. Epidemiology, it is now seen, is necessary to the solution of an array of pressing biomedical problems besides epidemic infections. These problems are mainly those of the causes of chronic disease, but they extend into such fields as the establishment of population norms of all kinds, the control of fertility, and genetics in populations.

The problems have in common only their potential origin in the environment, and the need for population studies to determine their origin.

Contemporary discoveries—the role of rubella in congenital anomalies; of cigarette smoking in lung cancer, emphysema, and heart disease; of oxygen in retrolental fibroplasia in the premature baby; of sex hormones in cancer of the breast and of the endometrium; and of viruses in Burkitt's lymphoma, hepatitis B, and kuru—arise from and are encompassed by a theoretic viewpoint sharply different from that of the early years of this century and the late nineteenth century.

The exemplary research of the earlier period dealt with infectious disease. Many infectious agents and modes of transmission were discovered. They were typically bacteria in acute disorders such as the major epidemic diseases, parasites in the tropical diseases, and later viruses. Yet most current teaching of epidemiology, in searching for relevance in the past, turns to quite another set of exemplars. What seems most meaningful are the discoveries that repeated ingestion of lead in small quantities caused the "Devonshire colic," that soot was a carcinogen, that measles was a transmissible if unidentified infection with a regular incubation period, that cholera was also caused by an unidentified infectious agent carried by sewage-contaminated water, that silent infection by an unknown organism underlay the mysterious pattern of transmission of poliomyelitis, or that pellagra signaled a nutritional deficiency.

This reordering of relevance is an unequivocal indication that epidemiologic thinking is guided by a new paradigm. The most salient methods are perceived to be those that can aid in the discovery of multiple environmental agents of unknown character or of obscure operation. Historical examples that conform with the new criteria of relevance are selected, both in the following text and in others, to indicate
the origins of various methods and the priority of those who invented them. An illusion of continuous linear development is created by this selective process. The resurrection of old priorities masks a leap from one paradigm to another (8).

The literature of epidemiology has come to show consensus in definitions of the discipline. The contention and debate about the scope of epidemiology to be found in the first two decades after World War II have disappeared. This institutionalization of the new scope and the new paradigm is a result of many factors. A major factor was the unprecedented growth of biomedical research sponsored by the National Institutes of Health (converted from the Hygiene Laboratories of the Public Health Service). The Institutes were staffed by a number of officers of the commissioned corps of the Public Health Service. Several rose to places high enough to affect policy. Aware of public health problems and needs, they could direct a share of funds and effort toward epidemiology, both for research and training. It was evident that before research could be produced, there must be persons on hand capable of producing it.

Thus, epidemiology, like other biomedical sciences, benefited from the training programs that were set up with such foresight, and on which much of the accomplishment in the United States in these fields came to rest. Few foresaw, before the late 1960s, any limit to the federal bounty. Many researchers, both teachers and students, were attracted from abroad. Graduate programs in epidemiology gradually began to take shape and to shift their focus from the magisterial to the doctoral level.

Another factor was the development of the Communicable Disease Center (now the Centers for Disease Control). Its Epidemic Intelligence Service (EIS) under Alexander Langmuir recruited, trained, and inducted into epidemiology (as an alternative to the military draft) many young physicians. In academic departments of medicine in the United States, there can be few infectious disease divisions without a former EIS officer on the staff. Departments of medicine are the flagships of their medical schools, and these physicians thereby contributed to the change in the medical climate.

In no small part, however, the intellectual force of the new paradigm was established by two major successes in research. One is the Framingham cohort study of heart disease; the second is the linking by case-control studies of lung cancer and cigarette smoking. As noted above, the adoption of the new paradigm entailed not only a shift in the object of study—from the agents of infectious disease toward the environmental causes of chronic disease—but also a shift in methods and techniques.

**The Systematization of the Literature**

The ideas initiating this phase of epidemiology are well reflected in a 1952 review of the discipline by John E. Gordon, Professor of Preventive Medicine and Epidemiology at the Harvard School of Public Health (13). In pointing to past triumphs, Gordon had to rely largely on research in infectious diseases such as yellow fever, mumps, influenza, infectious hepatitis, and pneumonia. In looking to the future, by contrast, he pointed to studies of chronic disease. He espoused an holistic view of disease, framed by the parameters of agent, host, and environment.

Aware of the analytic distinction between levels of organization, Gordon, like others before him, specified the population group as the unit of study. The viewpoint he adopted, in favor of a concept of multiple causality, marks a break from the pursuit of the specific agent initiated by the germ theory. The logic of this concept of multiple causes, one may infer, led Gordon to require the better definition of variables, especially the hypothetic causal variable. For this purpose, he turned to the social sciences and genetics in the hope that they would ad-
vance techniques for measuring the social components of causality.

The need for quantification at once follows from a population model of complex causes, and Gordon called for building on quantitative foundations in the manner of Wade Hampton Frost. It was already clear to Gordon that to undertake such an enterprise, epidemiologic research in the future would have to be a team affair that drew on several disciplines (social, biologic, and statistical) and on the skills of different specialists (especially clinicians and laboratory workers).

Gordon showed a keen appreciation of new directions of study and of the future focus on noninfectious disease. Aside from the classic work of Goldberger on pellagra (14), however, he could cite few results. As late as 1927, Frost, while wholly committed to a notion of epidemiology that encompassed the triad of agent, host, and environment, had confined the purview of the subject to specific infections. (It should be noted that he did this not so much on principle as to conform with contemporary usage and to permit more coherent discussion (9).) Gordon’s main examples of noninfectious disease epidemiology were from work in progress, or reviews, or conceptualizations of problems. Thus, he pointed to his own review of the epidemiology of accidents, his formulations with Harding le Riche of the epidemiology of nutrition, those of Theodore H. Ingalls of the epidemiology of developmental anomalies, and several others (dental caries, cancer, multiple sclerosis, heart disease, mental disorders, and the effects of air pollution).

In the little space Gordon devoted to research design, he differentiated the broad categories of descriptive and analytic epidemiology, and indeed he traced the historical evolution of epidemiology from the descriptive studies of John Graunt in 1662 through the comparative historical-geographic ecologic studies of epidemics of the 19th century (Finke, 1792; Hirsch, 1883, 1885, 1886) to the analytic studies of Peter Ludwig Panum (1847) on measles in the Faroe Islands, and of John Snow (1855) on cholera in London. He confined experimental epidemiology to the conduct of large-scale animal experiments, in a usage that persisted through much of the 1950s, and that had taken origin with the work by Amoss and Webster in the United States and by Greenwood and Wilson in Britain on cycles of epidemic transmission (15-17).

More revealing are Gordon’s omissions in discussing methods. In matters of design, with the exception of one new development, he referred only to the “field survey.” He implies that this design is cross-sectional and elicits prevalence. The one new development noted by Gordon he termed, for want of a name, the “field review.” He used this term, by which he meant the prospective cohort study of incident disease, to describe the freshly launched Framingham Study of coronary heart disease.

In fact, one searches in vain in the standard texts of the decade after World War II for a description of the basic features of either cohort or case-control designs. These are not to be found in Leavell and Clark’s work published in 1953 (18) or in the eighth edition of Maxcy and Rosenau of 1956 (19). This hiatus in the literature occurs despite the fact that, according to the recollections of Morton Levin and Alexander Langmuir, Frost taught the techniques of prospective and retrospective studies when they were students at The Johns Hopkins University School of Hygiene in 1933-1934 and 1940, respectively. (A hint in support appears in Frost’s 1933 paper and in Maxcy’s editorial comment on this paper (20).) Not until the work of MacMahon et al. in 1960 (21) does a systematic treatment of study design appear. As a result, this book became a founding text for modern epidemiology.

In Britain, in 1955, the sixth edition of Hill’s limpid work (22) incorporated clinical trials but not other designs. In 1957, Taylor and Knowelden’s book (23) touched briefly on the structure of designs, and in 1959, Doll’s chapter in a multi-author text
gave an exposition (24). The 1957 book by Morris (12) was seminal, but it taught from application not method.

With the broadening sweep of epidemiology, needs arose for conceptualizing and applying sociologic variables (7, 25); these needs too were met in new volumes (26, 27). Around this time also, texts on methods special to the study of limited classes of disease which dealt with the structure of study design began to appear (28–30).

I choose the systematic text as a marker for the general acceptance of an approach, a signal of the coherent synthesis required of a mature discipline. These textual crystallizations of theory and method reflect more than ongoing research. As Kuhn (8) argues, they appear when a new discipline has established a prevailing paradigm. Scattered papers on design topics of course are to be found over the years as researchers ran against design problems. Methodological developments flowed from a succession of designs tried in the field—cross-sectional field survey, cohort, retrospective case-control, quasi-experimental, and experimental—in an intriguing dialectic evolution. I shall treat each in turn. (Omitted from my discussion is the secondary analysis of vital data and time series in the tradition founded by William Farr. New developments include cohort analysis, advances in life table techniques and in mathematical modeling.)

THE CROSS-SECTIONAL FIELD SURVEY

At the time of the epidemic transition, the cross-sectional field survey was the readily available tool for epidemiologists seeking to describe the distribution of chronic disease. As early as 1916, Sydenstricker had used this approach in the study initiated by Goldberger (31) of living conditions of seven cotton mill communities at high risk of pellagra. In December 1921, Sydenstricker, by then Chief of the Office of Statistical Investigations of the United States Public Health Service, began a continuous series of morbidity surveys in Hagerstown, Maryland (32). (Hagerstown has been effectively maintained in the present day by George Comstock as the base for an "epidemiologic laboratory" for community studies (33).) A variety of morbidity surveys continued through the 1930s; a later series on chronic disease, planned by Sydenstricker and Frost, was in the end executed (by Jean Downes and Selwyn Collins) from 1938 through 1943 in Baltimore and completed after both the originators had died (34).

Field surveys do not have a prominent place as a research approach to causation in present day epidemiology, and the search for causes is now the nub of the discipline. In this light, I may seem to give surveys undeserved attention in what follows. One reason is paradoxical. It is precisely the manifest weaknesses of the survey method in causal research that engendered much of the thinking and the methods today found in the kit of the epidemiologist. Another reason for attention is that the field survey is a progenitor of current designs. First, it is an ancestor of the case-control design. Second, in the form of the longitudinal or panel study elaborated by repeated cross-sectional surveys over time, the field survey is also one of the ancestors of the cohort design.

Developments of the field survey were intensified in two directions in the post World War II phase. The first led to the institutionalization of morbidity surveys on a national level. The second led to the focus on community surveys of mental disorder.

A national morbidity survey of 750,000 families conducted by the Public Health Service with aid from the Works Progress Administration in 1935–1936 was in the direct lineage of Sydenstricker’s earlier surveys. After the war, beginning in January 1950, the Commission on Chronic Illness set out to determine the frequencies to be expected for the various diagnostic categories of chronic disease in urban and rural areas. (The first director of the Commission, who planned these surveys, was a
medical epidemiologist, Morton Levin.) What was learned, reported in four volumes (35), was largely incorporated in the continuing National Health Survey repeated at intervals since 1956.

The National Health Survey has three elements: repeated sample surveys of self-reported illness; medical examinations of smaller samples of defined age groups; and the collection of data from medical care facilities. The morbidity surveys, which rely on self-reports of diagnosed illness and of disability (measured by bed-days and dysfunction in daily living), have lately been extended to include nutritional, fertility, and other data. Although these morbidity surveys are descriptive and do not in themselves constitute etiologic research, they do constitute a resource for the generation of hypotheses, for the estimation of sample sizes, and for testing hypotheses by secondary analysis.

The second line of development of the morbidity survey—in community studies of mental disorder—is probably its main residual use in research. A signal innovation in the epidemiology of mental disorders after World War II was the attempt to escape the confines of medical diagnosis. Surveys of mental disorder, which in the United States had begun in nineteenth-century New England with Edward Jarvis, virtually all relied on counting medically diagnosed cases, most usually those found in medical care facilities.

Two basic problems had emerged. A minority of persons with mental disorders entered medical care, and, even for the minority in care, the diagnosis was unreliable between different psychiatrists and different institutions. Psychiatric epidemiologists therefore set about the task of forging instruments for sample surveys that would yield reliable measures of the prevalence of the slippery entities of mental dysfunction. The chosen instrument for breaking out of the constraints of treated disorder was the structured interview questionnaire—mainly in the form of symptom inventories—applied in community prevalence studies (36, 37).

The task was a long one and indeed is not yet completed. Two major surveys were conducted in the 1950s, the Midtown Manhattan Study (38) and the Stirling County studies in rural and small town Nova Scotia (39). The results of these studies were loaded with provocation and uncertainty. The reported frequencies of "mental impairment" or psychiatric dysfunction were met with open disbelief because of their magnitude. Validation was not attainable because the scores obtained on the symptom inventories could not be extrapolated to the disorders diagnosed by psychiatrists. Psychiatric researchers, aware of the problem, tried at least to specify "caseness," a condition a psychiatrist would judge in need of treatment (39).

To bridge the gap and define the difference between symptom inventories and psychiatric diagnosis, the attempt is being made to create, and to test in the field, psychiatric screening and diagnostic instruments that relate to the standard diagnostic manuals used by psychiatrists. In other areas of research, also, many scales of function and disability have been invented which are specialized, standardized, and to some extent validated. Some such scales relate to nonpsychiatric disorders, for example respiratory disease, and others relate to life stress and to life situations.

The labor devoted to enhancing the strength of cross-sectional surveys has not expunged the weaknesses in the reliability, validity, and significance of their measures. While the weaknesses remain, the efforts to correct them have encouraged a rigor in method that has been to the advantage of epidemiology in general. Two areas in particular have benefited. One is the design and analysis of surveys; the other is logical inference. These benefits can best be understood in context.

To take first survey design and analysis, most epidemiologists are intensely concerned with the gathering of data and with
ensuring that it meets minimum standards of quality. The methodologists who have treated systematically such topics as the construction of scales, interviewing and response bias, reliability, validity, and measurement error have been drawn, oddly, not from epidemiologists but from psychologists and sociologists. These advances, acquired from the social and psychologic sciences (6, 36), were the fruit of the period of the dominance of functionalism and positivism in those fields, a period during which they strove to emulate, in the study of society, the rigor and methods of the physical sciences.

At the same time, medicine was obliged to recognize that it was not self-sufficient. In many medical research areas, reliance on nonmedical scientists was growing as the research became ever more specialized. In the major surveys of the Commission on Chronic Illness (35), or Midtown Manhattan (38), or Nova Scotia (39), sociologists and psychologists were recruited as the technicians and the architects of design. In the process, a new breed grew up of social and psychosocial epidemiologists; only a scattering of such medical epidemiologists as Sidney Cobb and John Cassel had preceded them. In short, these various studies opened the way for the substantial development of psychosocial epidemiology in the United States. They attracted a host of new researchers and generated a stock of new and better methods.

This accession of social scientists was important also to future recruitment to graduate studies in epidemiology. After the first flush of enthusiasm for social issues in the ambience of World War II had subsided, the long labors and meager material rewards of epidemiology and public health seemed the less enticing beside the quick returns and the goldfields of medicine, and talented recruits were difficult to find. The new social epidemiologists filled the gap. Their presence and productivity demonstrated the potential of epigones untrained in medicine for carrying the work of epidemiology forward.

In any event, the texts on survey method of this period are written by sociologists, psychologists, and statisticians. These include such special topics as sampling theory (40, 41) and interviewing (42) as well as general texts on survey research (43, 44). Much stemmed from the writings of Fisher (45, 46). It was 1959 before a text—a British one—was devoted to health and medical surveys (47). A single-author text on the subject, from Israel, did not appear until 1974 (48). I am not aware of even a simple epidemiologic text on interviewing until 1975 (49). Again, this text was British.

Perhaps traditional epidemiologists dwelt little on these matters because, on the face of it, the data they sought treated objective manifestations with serious implications for affected persons. Did such data seem more grounded in material reality—and therefore less in need of being challenged for truth and tested for error—than such things as subjective opinions and the tenebrous recesses of states of mind?

Again, perhaps it was because epidemiologists dealt with categoric data produced by the diagnostic process that they made their chief methodological contribution to cross-sectional studies. This was to recognize and measure the unreliability of diagnostic observations. Thus, their concerns were with intra- and interobserver error; with misclassification and its consequences for analysis and inference; and with instrument reliability, as in the comparison of self-reported illness with medically assigned diagnoses.

One field, the early detection of disease by diagnostic screening, epidemiology made its own. The field arose in response to the demands of the public health strategy of secondary prevention through early detection and treatment of disease. Its theory was built around the fourfold table, which is at the center of modern epidemiologic thinking. Screening theory has since been taken up in clinical pathology and other
diagnostic fields, as well as in improving the efficiency of surveys. Thus, the technical operations of screening include measures of sensitivity, specificity, predictive value, conditional probabilities, lead time bias, etc. (50–53). The synthesis of screening theory—in a comprehensive model which reconciles these notions with those of reliability and validity on the one hand, and with those of decision theory on the other—is still to be achieved.

Now let us consider what benefits field surveys have conferred on logical inference and interpretation. These problems, first taken up systematically by philosophers of science and statisticians, were an early preoccupation of survey researchers, a preoccupation that is unsurprising in face of the ambiguity and instability of social research data. In epidemiology, the concern with logical inference reflects the shift of focus from the discovery of infectious agents toward etiologic studies of chronic disease. Early attempts to apply the then dominant means of the cross-sectional field survey to the ends of chronic disease studies made the problems of inference at once obvious.

Yet again, the relevant literature on logical inference from observational data was not indigenous to epidemiology but was largely the product of the social sciences. A substantial body of literature flowed from the practical demands for surveys made on American social scientists during World War II. Information was wanted about people at home, in the armed forces, and in occupied territories (54). These social scientists had to contend with two problems of inference exaggerated by cross-sectional data. Such data lack bench marks for time order among variables, and they run a high risk of confounding causal relationships in a proliferation of variables.

Statisticians helped to cut paths through the multivariate undergrowth, as did those of other disciplines wrestling with similar problems. This concern with confounding and inference led to the early adoption by the social sciences and psychology of the methods of multiple variable analysis invented by Fisher (45, 46) and others. Social surveys often used continuous variables that lent themselves easily to multiple regression analysis. Once computers made such analyses accessible to anyone who could command a packaged program, they also lent themselves, alas, to mindless abuse. Multivariate analyses that neglect the known and unknown relationships and causal pathways among variables are as likely to obscure as to reveal reality.

The advance of chaos was stemmed somewhat with the reintroduction of path analysis to the biologic world. The method was invented around the time of World War I by Wright (55), a founding father of population genetics. Path analysis, however, was little used until taken up by the new discipline of econometrics, and then by social scientists, finally to return to population genetics and the biomedical world in the 1970s. Path analytic models impose a degree of order because the analyst must commit himself to explicit causal models. Hence, assumptions are more readily brought to the surface and made visible. Further developments in this direction take account of reciprocal pathways (structural equations) and longitudinal relationships through time (time lagging and causal analysis).

In epidemiology, we find the first treatment of the problems of logical inference within a multiple cause framework in the text of MacMahon et al. in 1960 (21). These problems have since been extended and elaborated on in other texts, beginning with a concept of variables located in a structure of causal sequences, and with a set of strategies to establish what the sequences might be and to screen out extraneous variables (6).

A substantial literature on confounding, interaction, and various estimates of risk has now developed. Epidemiologists began to make general use of multiple variable analysis only after the invention of the
binary regression method (the use of dummy variables by which discontinuous variables could be entered into multiple regression equations) and logistic regression (56–58). One may anticipate increasing use and new borrowings of multivariate analytic approaches.

In summary, the attention given by social scientists to the inherent weaknesses of field surveys led to the strengthening of design techniques and the creation of methods for constructing and testing instruments and scales and for dealing with measurement problems. It also led to awareness of the demands of logical inference and the dangers of confounding and, hence, to the early adoption of multiple variable analysis. Because of their training and skills in these matters, social scientists have become a growing segment of the epidemiologic profession.

COHORT STUDIES

Among doctors in the United States, the Framingham Study stands as the exemplar of epidemiology. It is the epitome of successful epidemiologic research, productive of insights and applications such as cardiovascular risk factor scores. Among epidemiologists in the United States, the Framingham Study has become the prototype and model of the cohort study. For these reasons, it deserves close examination.

As best one can determine from accounts of the history of this study—like all histories, these are not in perfect agreement (59, 60)—the Framingham Study set out to develop case-finding procedures for heart disease in healthy volunteers. The seeming intention was to establish incidence rates in a general population.

Joseph Mountin, assistant to the Surgeon General, United States Public Health Service, saw the public health interest in the rising epidemic of heart disease. With Mountin's support, David Rutstein, Chairman of the Department of Preventive Medicine at Harvard Medical School, joined with the Massachusetts Commissioner of Health to have a follow-up study set up at Framingham, a town of 28,000 people outside Boston. The study was initiated in October 1947; medical examinations of volunteers began a year later (October 1948). Well into the second year (July 1, 1949), the newly founded National Heart Institute took over the study. Only then did it emerge that the aim of the study would be to determine "factors influencing the development of heart disease" (59).

A new, more representative sampling of 30- to 59-year-old persons was then planned, and a required sample size was estimated (there is a hint in one account that in actuality the estimate made was not of required sample size but of the feasibility of the study within a fixed limit of 6,000 subjects). A revised protocol for biennial examinations over a 20-year period was put into operation by a new principal investigator (Thomas Dawber) some 15 or 18 months after the first cycle of examinations had begun. Only about nine months later than that, at the end of the third year (October 1950), was a laboratory available to carry out cholesterol and other determinations. Smoking histories were not taken at first and, once introduced, were modified. About two thirds of the selected sample agreed to participate, and thus numbers were supplemented by volunteers (available because it was judged unwise to turn away anyone desiring an examination).

As the study progressed, instruments were added and refined to improve the quality of measurement and to test additional risk factors such as physical activity, diet, and life stress. Because of the proportion of refusals, the study could not yield the sound representative incidence rates intended. One suspects also that the distinction between cumulative incidence in a cohort of constant composition and period incidence rates in a general population with continuous accessions and losses had not generally been appreciated, despite Farr's clarification in 1859 (61, 62). Perhaps it was with this realization that the goal of
establishing the individual risks attached to specific factors evolved. This rationale for longitudinal studies, now routinely assumed, seems not to have been clear and certainly was not explicit at the outset.

When it came to analysis, tactics had to be devised to deal with the variability of repeated measurements over time. Where there were unreliability and fluctuation in values, how much was the result of regression to the mean? and what were the appropriate base lines for starting measures against which to assess change (63)? These were major difficulties inherent in longitudinal observations; the continuity of the project gave some opportunity for solving them. Two of the investigators, in describing the history of this study, wrote that it “might be reasonable to consider measurement problems as one of the justifications for longitudinal studies” (59).

Finally, as the study continued, it became apparent that there was no good way of stopping it. (At a meeting of the Royal Society of Medicine in London held to celebrate his sixty-fifth anniversary, Sir Austin Bradford Hill read a paper on the 10-year anniversary of the national cohort study of the effects of smoking among British doctors. He opened with the remark that there was nothing to be done with a longitudinal study but to pass it on to someone else in one's will.) Even now no agreed stopping rules for a longitudinal study exist.

After 20 years, the National Heart Institute decided that the Framingham Study had realized its original goals, and funding was withdrawn. Determined supporters of the study obtained funding elsewhere, however. After an interval of a few years, the project was once again funded by the National Institutes of Health. The project is now devoted to stroke, cerebrovascular disease, and various aspects of aging, and also to follow-up studies of the offspring of the original participants.

I have discovered no true and complete precursors in epidemiology of the Framingham longitudinal study. Conceivably, a number of longitudinal growth and development studies started in the 1920s and 1930s—"auxological epidemiology"—could be considered in this light (64-66). Also, Sir James MacKenzie, founder of modern cardiology, had sowed an idea that can be followed to Framingham.

Around 1920, MacKenzie had left laurels and fame in London to set up the St. Andrew’s Institute of Clinical Research for the longitudinal study of the natural history of disease in the community through the agency of general practitioners in Dundee, in his native Scotland. His aim was to discover the physiologic expressions of disease, from the "beginnings of illness" and throughout its course. Although MacKenzie died in 1925 before his plan could be fully realized, the germ of his idea was carried back to Boston by Paul Dudley White. White had gone to study with MacKenzie and had in turn become a leading cardiologist of his time and an early protagonist of the Framingham idea. The modern techniques of the longitudinal study, however, were still to be invented and its problems discovered.

One might argue that in principle the cohort design is an analogue of two preexisting types of design that are forward looking and involve the lapse of time. These are the population study of incidence and, as mentioned above, the repeated cross-sectional survey. A pointer to the connection of the cohort design with studies of incidence is the original intention of Framingham to determine prospectively, not the individual risks of heart disease, but its incidence in the population at large. The connection with the repeated cross-sectional survey rests on the fact that such surveys afford a means of determining prospectively changes in state. In sociology, when applied to a fixed population sample, such surveys constitute the so-called panel design; except for the substance under study, they do not differ from cohort studies that yield both cumulative incidence and
probabilities according to individual characteristics.

Yet, the crystallization of the cohort method must be attributed to the requirements of studies of chronic disease. This impetus is already seen in Frost's 1933 paper on the risks of familial contact in tuberculosis (20). Tuberculosis was a bridging condition for epidemiologists making the transition from the study of acute infections to chronic disease. (With respect to the bridging role of tuberculosis, I shall have occasion to cite later the development of follow-up studies of reconstructed cohorts. In addition, the first application of cohort analysis to disease was Andvord's analysis of tuberculosis mortality. The innovative multicenter clinical trial for the evaluation of streptomycin in treating tuberculosis and the early prophylactic trials of Bacillus Calmette-Guérin (BCG) also reinforce this notion of a bridging function.)

The Framingham Study, in turn, has been a stimulus of three lines of development, namely, 1) other cohort studies of cardiovascular disease; 2) other types of cohort studies; and 3) long-term community studies.

In the first group, an array of cohort studies of cardiovascular disease in communities and in large organizations have tested, validated, and extended the Framingham results. The earlier community cohorts include Tecumseh, Michigan (67); Evans County, Georgia (68); and Honolulu (69). Cohorts of employees were followed in Minneapolis (70); Albany, New York (71); Los Angeles (72); and Chicago (73, 74).

In the second group, a number of other and less famous longitudinal studies, most often of particular exposures and multiple diseases, evolved in parallel with the Framingham Study. This strategy, of following multiple outcomes in a cohort exposed to a specified risk, is in fact better fitted to the logical uses of longitudinal observation than the initial Framingham approach which considered multiple risk factors but strictly limited the outcomes.

To the testing and cross-fertilization among all these various studies we owe our understanding of the longitudinal method. Among the better known in the United States are the continuing follow-up of the Japanese survivors of atomic bombs (75); several follow-up studies of United States veterans (76) and notably Dorn's study of smoking and mortality (77); the American Cancer Society study of the effects of smoking (78); and the Collaborative Perinatal Project on antenatal factors in postnatal child development (79).

The third group—long-term community studies—took up the thread spun on the one side from Sir James MacKenzie (80) and on the other from Sydenstricker's repeated cross-sectional sample surveys of Hagerstown, Maryland (32). From these evolved the idea of the community as epidemiologic laboratory (81, 82). Such parochial studies have in common the continuing observation through time of a defined community. Their object may be accomplished by follow-up (Tecumseh, Michigan (83); Human Population Laboratory, Oakland, California (84)); or by repeated cross-sectional survey or census (Washington Heights, New York City (85); Washington County, Maryland (33)); or by registers of a given disorder (mental disorder in Monroe County, New York (86)); or by linked record systems (Olmsted County, Mayo Clinic, Rochester, Minnesota (87)).

Judging by published results, most community laboratories have produced less than was hoped for, and perhaps less than their potential. Aside from the stopping rule problem, small communities are not likely to yield adequate numbers either for hypothesis testing or for the study of secular trends unless the disorders under study are very common. In general, small communities serve best for monitoring trends of incidence and prevalence and as a frame for drawing cohort or case-control study samples.
The use of a special community census for sampling in multiple studies is well illustrated by a continuous stream of research reports—on such diverse topics as the effects of the urban factor in lung function, of widowhood on mortality in general, and of soft water on cardiovascular disease—by the Washington County Epidemiological Laboratory (33, 88-90). In this way, community laboratories can also serve to test a well grounded hypothesis. For instance, a seminal test of the effect of social support on mortality was conducted by the Oakland Human Population Laboratory (84). The long-standing record-linked system of the Mayo Clinic has proved highly productive in each of these areas (87). This system has provided the best available frequency and prognostic data on epilepsy, stroke, and many other disorders. It has also been quickly deployed to retest many hypotheses, such as the effects of reserpine on breast cancer, and of diethylstilbestrol on clear-cell adenocarcinoma of the female genital tract.

By the 10-year mark of the Framingham Study, enough had been learned about cohort studies for a full exposition to appear in MacMahon et al.'s 1960 text (21). Doll (24), in 1959, recommended limiting their use primarily to supporting or refuting a well formulated and precise hypothesis. Noted in favor of the method were the lesser opportunity for bias—since knowledge of the outcome could not affect the estimates of exposure—and the simplicity of the relative risk estimates. One may add the advantages of the design for the study of rare and extreme or high-dose exposures in definable groups, as demonstrated with mustard gas and World War I veterans (76), with the atomic bombs in Hiroshima and Nagasaki (75), or with asbestos dust in New York construction workers (91).

If the exposure or risk is common, advantage may lie in the detailed description of the factor that can be gathered concurrently with the field observations, as in Framingham. In that study, moreover, repeated follow-up allowed both risk factor and outcome data to be extended and refined. This flexibility overcomes the handicap of long-term cohort studies with a once-only starting point which are tied to initial hypotheses. By the time the results of such studies are known, new knowledge may have made their hypotheses anachronistic.

 Whatever the nature of the exposure, multiple and unknown outcomes of a given exposure are not easily discovered except by means of the cohort study. One advantage unique to the design is that only longitudinal observation can take account of selective loss and survival in the time interval between exposure and outcome. Where these losses can occur as a result of the exposure, they may cause serious confounding.

Thus, in the populations exposed to the atomic bombs of 1945, effects of prenatal exposure on measured intelligence were found only for exposures after the sixth week of gestation. Early loss of damaged fetuses probably suppressed the evidence of effects of earlier exposure. This attrition is indicated by a deficiency in the number of births among the cohorts exposed early in gestation (92). Likewise, in a study of the effects of the Dutch famine of 1944-1945 on human development, a reduction in the fertility of the poorer social classes as a result of famine so distorted the social composition of birth cohorts as to give the appearance that severe famine exposure early in pregnancy improved the mental performance of young adults (93).

Various economies have been devised to enhance the advantages of cohort studies. An historical cohort study makes use of data previously collected to define cohorts and outcomes, as Case et al. (94) did in a study in Britain of bladder cancer in chemical workers. An American example was a study of the effect on mortality of occupational exposure of medical specialists to ionizing radiation (95). A later example was the study referred to above of the Dutch
famine of 1944 to 1945; for birth cohorts defined by exposure, outcomes were studied at birth, in terms of mortality into adulthood, and in young adult males at military induction (93).

Economies have also been achieved where the exposed cohorts could be constructed from already collected data, but where outcomes were determined at concurrent follow-up. This approach was pioneered in a study of the mortality of patients after discharge from the Adirondack Cottage Sanitarium, reported in 1904 by Brown, a tuberculosis specialist, and Pope, an actuary then a patient in the sanitarium (96). The attempt ran foul of a methodological difficulty. Although the observation period began with the onset of symptoms, the cohort comprised only individuals who had survived to admission or discharge and could not account for selective bias by attrition up to that time.

In 1933, Frost noted, with his usual unnerving concision, his application of a variant of this method to the familial spread of tuberculosis (20). Subsequent studies of secondarily constructed cohorts, such as those of the Atomic Bomb Casualty Commission (75) and the US veterans (76), were eminently successful. With data gathered at follow-up, this type of study obtains the advantage, at some extra cost, of meeting the precise needs for measurements of outcome data instead of having to rely on what opportunity brings.

Almost in passing, Frost carried the method a step further in order to study the aggregation of tuberculosis in families. To do this, he reconstructed the mortality experience of family cohorts separately around cases of tuberculosis and around unaffected controls—which is to say, a historical cohort design is constructed around subjects selected in the manner of a case-control study (97).

The Framingham Study itself has left its mark. This is so much the case that among many physicians and referees for journals, the conviction persists that only prospective studies can yield sound epidemiologic knowledge. Yet today, if one is to judge by the initial protocols referred to in various accounts of its history (59, 60), the Framingham Study could not hope to be funded. Peer reviewers would demand a sound rationale for the study; structured hypotheses; an estimate of sample size likely to yield significant answers (derived from the levels of significance and power, anticipated relative risk, and frequency of the outcome in the unexposed population); a description of the mode of recruitment of subjects, including criteria for eligibility; details of how the data are to be collected; the type and quality of the data and measures of their reliability and validity; and in addition, an account of how the data are to be analyzed.

A matter for historical and philosophic reflection is whether the standards we have created and rightly apply are not in some way lacking. For the Framingham Study was indeed one of two major intellectual levers in bringing about a shift in the chronic disease paradigm, and it is undisputedly the foundation stone for current ideas about risk factors in general and the prevention of ischemic heart disease in particular.

CASE-CONTROL STUDIES

Cohort studies and community studies are expensive. Case-control studies are less so and have come to dominate American epidemiology, for not only can they examine specific hypotheses but they also function as exploratory studies. Cole (98) reports that in 20 years (since the mid-1950s), case-control studies published in four leading journals increased four- to sevenfold.
Sartwell evidently introduced the term (99); MacMahon et al. (21) used the term case history to escape the ambiguities of the term retrospective; Miettinen (100) has used the term case-referent; and Hulka et al. (101) have used the term case-compeer.

In conformity with my remarks about the effect of the epidemiologic transition on epidemiologic methods (and as Cole argues), the proliferation of case-control studies is the consequence of the predominance of chronic disease and of the utility of this method for studying disease with a long latent interval between exposure and onset. The backward looking study collapses the time interval between exposure and outcome and avoids the labor and time involved in awaiting the slow evolution characteristic of chronic disease. Equally to the point, the case-control method, unlike the cohort method, is suited to the investigation of small clusters of chronic disease that appear sporadically.

The method has proved efficacious too, if within limits (102), in the study of epidemic outbreaks of acute disorders of unknown origin, as with such newly identified infectious disease syndromes as Legionnaire's disease (103) and toxic shock syndrome (102). The design has served no less well for chronic disease caused by infection with a long latent period. Infections have assumed fresh significance since the discovery of the slow viruses of kuru and Creutzfeld-Jacob disease, the identification of hepatitis B virus as an agent in both "serum jaundice" and hepatic cancer, and the emergence of syndromes with an immunologic basis such as subacute sclerosing panencephalitis and acquired immunodeficiency syndrome. This turn of the wheel, toward the study of infections by means of research designs typically used for chronic disease, points to an essential if seldom manifested unity of infectious and noninfectious disease epidemiology.

In principle, as mentioned above, the case-control study can be seen as an analogue of the prevalence survey of morbidity, a development from it that vastly increased efficiency. Both approaches begin from the manifestation of morbidity; both proceed to acquire information about the attributes of subjects of the survey sample, so that those who exhibit morbidity can be compared with those who do not. Like the field survey, the case-control design is cross-sectional in that current attributes and exposures, or a current history of past attributes and exposures, are compared. The attention now given routinely to retrospective—so that past history is documented with maximum precision and the risk of confusing the time order of cause and effect is reduced—is a subsequent if natural evolution of the approach.

The link between the two designs is emphasized by difficulties in separating them. Studies that compare the attributes of cases with those of a simultaneously observed population at large from which the cases are drawn are clearly cross-sectional field surveys. When the comparison of cases is with a sample of the population, however, ambiguity enters into classification. Somewhat arbitrarily, if the survey aims to establish the relationships of cases to antecedent factors in a retrospective manner, I assign it to the case-control category.

The use of a cross-sectional approach that resembles the case-control design has been traced in papers of William Guy in mid-nineteenth century England (104). A well developed form of the method, used by J. E. Lane-Claypon in 1926, is advanced by Cole as the first “true case-control study of modern type” (101). The prototypes that initiated the present upsurge in case-control studies, however, are those in 1950 on smoking and lung cancer.

Although a pre-war case-control study by Müller (105) in Germany strongly supported the association of smoking with lung cancer, conflicting results prevailed in the literature up to 1950. In that year, four successive reports of case-control studies showed an association (106–109). (Schreck et al. (106) acknowledge the stimulus of an
earlier verbal presentation of results by Wynder and Graham. From an account by Levin (personal communication, 1984), the fact that the papers of Wynder and Graham (107) and of Levin et al. (108) appeared in the same issue of the Journal of the American Medical Association was due to a happy accident. The hypothesis of the association between smoking and lung cancer was not well known and hitherto undocumented, so that the editor was reluctant to publish these papers. Dr. Levin, however, was subsequently able to convince the editor of the importance of both papers, which were then published in the same issue.)

The persuasiveness of the last of these, Doll and Hill’s study (109), was enhanced by the consistency of its results with those of its predecessors. In its own right, however, this British paper set a new standard for the case-control study. This work—together with Doll and Hill’s national cohort study of British doctors that followed (110)—provided the second major intellectual lever in shifting the paradigm of chronic disease from an intrinsic to an environmental causal model. It stands as a classic exemplar for the investigation of a given outcome and an array of exposures. No previous research paper lays out the essentials of the case-control method with such understanding and meticulous care.

The authors consider a wide range of design problems. Their analysis deals with the comparability of cases and controls and the precision of matching (for age, sex, place of interview); selective recruitment and attrition; confounding (by social class); the validity of the outcome variable and ways of refining it; the quality of the hypothetic causal variable, including the reliability of the history of exposure, the persistence of recall, and interviewer and response bias; measures of the hypothetic causal variable, including refinements of dose (amount and duration) and of type (variations in kinds and methods of smoking).

In their interpretation, the authors raise alternative hypotheses and dispose of alternative explanations that could reside in faulty study design or execution. The hypothesis that smoking is causal is reviewed in the light of its consistency with other results and its coherence both with population data on the distribution of lung cancer and smoking and with animal experiments.

Finally, Doll and Hill (109) set out statistics for the significance levels and strength of the observed associations. Noteworthy is their extreme caution when they make estimates of relative risks (not odds ratios). Thus, the authors lay out their assumptions in detail. Their caution extends to not actually stating the risks, although these are on display in a figure. This tentativeness in estimating risks underlines the importance of the subsequent contribution of Cornfield (111) and the developments that followed from it. In this 1951 paper, Cornfield established today’s everyday knowledge of the conditions under which a simple odds or cross-products ratio fairly represents relative risk.

Epidemiologists had to wait until 1959, however, before Mantel and Haenszel (112) gave us the procedure that lifted the case-control study into the multivariate age. The procedure, the basis for which had been laid by Cochran (113), permitted the weighted combination of chi-square differences for the odds ratios of stratified variables. In 1973, Fleiss (114) constructed his statistics text around the analysis of rates and proportions; this was the first text to meet the particular needs of epidemiologists in analyzing the fourfold table.

One subsequent current of thought has sifted the assumptions for attaining comparability between cases and controls, especially matching. It is fair to attribute a good deal of development to the poignard of Berkson’s skepticism of the case-control method (115). In the United States, Miettinen (116) has been a leading contributor to our understanding of the utilities and inutilities of matching, as well as to
problems of confounding and risk estimation. We have learned that in most instances, excepting either known relationships or extreme confounding, the procedure is best reduced to a minimum. Very recently, we have a text wholly devoted to the exegesis of case-control studies (117) and another largely devoted to the multivariate analysis of such studies by logistic regression and log-linear methods (118).

Furthermore, the versatility of case-control studies has been greatly enhanced by the advent of computer programs that permit the ready use of complicated methods. Log-linear analysis and its offshoots address directly the use of multiple categoric variables, without at once running the analyst into the problem of diminishing numbers in each cell as the data are stretched thin across stratified contingency tables (119–121). By mathematical modeling, statistical virtuosi can reduce complex associations adjusted for multiple variables to simple and elegant curves. As always, in interpreting such masterly simplifications, their assumptions must be closely watched.

These are recent developments. The utility, the robustness, and the hazards of the case-control design have gradually come to be appreciated. Thus, several tests on cohort data—including the Framingham data—have shown that the case-control design yields essentially the same results. Although there has been a steady increment in the number of case-control studies through the decades of the 1950s and the 1960s, many fields remain unexplored by this technique. To take one example, Lilienfeld and Pasamanick (122) reported a case-control study of childhood epilepsy in Baltimore in 1954. Aside from a related replication in the same city among blacks only, by Henderson et al. (123), the findings remained untested by any case-control study of epilepsy over the next 25 years. (In this instance, the findings were in fact not replicated in cohort studies (124).) Essentially the same applies to the case-control study of cerebral palsy in New York State by Lilienfeld and Parkhurst in 1951 (125).

The use of the case-control design has been extended in various ways. One extension is to economize in the use of large-scale longitudinal cohort data. Thus, in the "nested case-control study," an outcome of interest selected from the cohort data comprises the cases. The recorded antecedent experience of the cases is compared with that of a sample of unaffected individuals drawn from the cohort to serve as controls.

Another extension is in the monitoring of population hazards. Thus, the Boston Collaborative Drug Surveillance Program (126) set out to monitor adverse drug effects by continuously sampling patients admitted to hospital wards and taking their histories. Initially, the monitoring related to current hospital experience, but it was soon expanded to relate to drug use over the months previous to the admission diagnosis. The matrix of data on drug exposure and diagnoses was systematically screened for associations by computer. Associations were examined in detail for robustness and then followed up and tested in case-control studies. In an early example, a relationship of thrombophlebitis to oral contraceptive use conditional on blood group was demonstrated.

This approach has not fully conquered the problem, common in epidemiology, of assessing the statistical significance of associations of multiple factors with multiple end points. In general, in a particular instance, the best resort is to the logic of causal inference as the guide to judging statistical inference. Ultimately, only consistency on replication under varying conditions can fully reassure.

A different approach to monitoring is used in a continuing case-control study of spontaneous abortions (127). The cumulative summation (CUSUM) method has been applied in Britain to monitor time clusters of congenital anomalies (128) and was here adapted to monitor chromosomal anomalies. The exploration of such clusters
through special case-control studies, however, has not yet proved itself productive.

The case-control method, we may conclude, is now in full bloom. Refinements of analytic techniques to control confounding and to achieve comparability cascade from the journals. The method is beginning to be applied in evaluation research, such as the efficacy of screening and secondary prevention (129), of prophylaxis for primary prevention (130, 131), and of medical care. Further extensions and uses are bound to appear.

Epidemiologic Quasi-Experiments

The quasi-experiment has a long history in medicine and public health. Typical experiments test the effects of specified interventions in a predesignated group by comparison with other predesignated groups similar in all respects but the intervention. Quasi-experiments likewise test interventions, but either an experimental or a comparison group or both are not predesignated.

Pierre Louis, the teacher of several founding figures of modern epidemiology including William Farr, reported an early quasi-experiment in Paris in 1835. This study survives as an historic episode of debunking. In this study, Louis examined the effects of bleeding by the application of leeches. Broussais, a great figure at the Sorbonne, was a chief advocate of bleeding, and millions of leeches were imported into Paris each year. Louis studied the outcome of pneumonia in patients admitted to hospital—and hence bled—at successively longer intervals after the onset of the illness. Fatality rates showed a distinct gradient, with the highest among those who were bled the soonest. The odds ratio, for those bled on days one to three of the illness compared with those bled on days seven to nine, was more than 5 (132).

This study is said to have ended the vogue of bleeding as a panacea. Louis was careful to note that the illnesses were equally severe in those admitted early as in those admitted late; to be just, however, one must note that in his study Louis could not control for selective survival. Patients who survived pneumonia long enough to be admitted late in the illness might very well have passed through the worst of the illness by the time they were bled.

In the field of public health, the quasi-experiment was early on display in Edwin Chadwick’s 1842 Report on the Sanitary Conditions of the Labouring Population in Great Britain, in which he compared mortality in various towns before and after the introduction of “drainage” (closed sewage systems), sometimes with comparable towns as controls. In another instance, Farr examined the effects of the “moral treatment” for mental disorder introduced by John Connolly at Hanwell Asylum in the 1830s. Farr showed the mortality rate to be only half as great at Hanwell as in comparable institutions that used different modes of care.

As these examples suggest, the quasi-experiment is an indispensable design for the evaluation of programs of intervention. Recent examples include an evaluation of the effect of fetal monitoring on the outcome of obstetric deliveries in Boston (133); an evaluation of the effect of newborn intensive care on newborn mortality over three years in New York City (134); and an evaluation of the efficacy of infection surveillance and control in the prevention of nosocomial infections in US hospitals (135).

This strategy and these results carry high risks of bias and confounding for the unwary. The abiding problem in all such studies, as with that of Louis on bleeding, is the potential for bias in the selective recruitment of those exposed to the intervention. The risk that both exposure to the treatment and the outcome are influenced by the attributes of the recruits is higher when they are volunteers.

Other problems are held in common with experimental and longitudinal studies. They include the inadequate specification
and demonstration of actual exposure to program or treatment input; the difficulty of separating the effect of the specified intervention from the activity that is the necessary accompaniment of administering it; selective dropout; and regression to the mean, inevitably present whenever there is measurement error—which is always. Quasi-experimental designs probably yield their strongest results where complete populations are observed, where input is well specified, and where a sufficiency of data on crucial confounding factors is available.

With this design, epidemiologists have once more had to turn for a systematic exposition to texts by methodologists from other disciplines, especially psychology. Notable among these are the works of Campbell and colleagues (136, 137).

**Epidemiologic Experiments**

Although the experimental design is certainly the least common type of study undertaken by epidemiologists in the United States today, it is far from the least important. The experiment is the design of choice where the investigator seeks internal validity in results—which will usually be where a highly specific hypothesis can be tested—and where precision is wanted in the estimate of effects. These conditions are characteristically sought with medical treatments, especially medicaments. It is consistent that the most famous of the early experimental trials, published in 1753 by the naval surgeon James Lind, demonstrated the superiority of lemons over several other treatments for scurvy.

The randomized trial, although descended from Ronald Fisher by way of Austin Bradford Hill, is not only, or strictly, epidemiologic ground. This ground, best defined by statisticians, has overlapped with that of clinicians, pharmacologists, and others concerned with therapy. Epidemiologists are perhaps more often called upon in multicenter trials, as with the University Group Diabetes Program trial of oral hypoglycemic agents, controversial for its demonstration of adverse effects only (138). Indeed, the hallmark for modern clinical trials is the multicenter trial of streptomycin in the treatment of tuberculosis, designed by Hill for the Medical Research Council in Britain in the late 1940s (139).

Epidemiologists are on their own firm ground, however, with prophylactic trials of vaccines, as in testing the BCG vaccine for tuberculosis (140). In the United States, the most dramatic of these was the landmark poliomyelitis trial of 1954 (141). The drama resulted from the threat to healthy children—especially those from the higher social classes—of a devastating epidemic disease. It was heightened by the high level of publicity generated by the voluntary agency (the March of Dimes) that funded the trial, and by concerns about the unknown risks of the killed Salk vaccine. The drama was further increased soon after by the poliomyelitis cases caused by contaminated lots of the Cutter vaccine.

This prophylactic trial involved nearly two million children. As initially designed, the trial was not randomized. It took much pressure from participating scientists to bring about a second randomized design, which was then superimposed and run parallel to the first. After the trial, the severity of the criticism of the nonrandom segment was such that a large-scale nonrandomized trial of a vaccine is unlikely soon to be lightly undertaken. This caution does not hold for other preventive interventions as many examples show (a recent one in England being the prophylaxis of neural tube defects by preconception multivitamin supplements, in which hospital boards blocked the randomized approach on allegedly ethical grounds (142)).

A prophylactic trial can avoid the necessity for very large numbers if a population can be found with a high incidence of the condition to be prevented. A trial of hepatitis B vaccine used this high-risk strategy to bring one cycle of the saga of hepatitis virus—at the center of which is Baruch
Blumberg's discovery of the Australia antigen—to a satisfying conclusion. Szmuness and colleagues (143) identified several groups at high risk of hepatitis B infection: people living under poor sanitary conditions in the Third World; renal dialysis patients; and promiscuous male homosexuals. In a masterly study, Szmuness recruited and randomized some 1,083 active male homosexuals in New York City within one year. Eight months later, he had given definitive proof of the vaccine's efficacy.

The case for randomized trials for the testing of drugs and vaccines seems unassailable, although there is room for trying out modifications to make things easier, for example sequential trials, factorial designs, lesser numbers of controls, and others. With regard to preventive experimental intervention, Joseph Goldberger, discoverer of the nutritional cause of pellagra, is a revered role model for American epidemiologists (14). His conclusive demonstration of a dietary deficiency in pellagra depended on dietary experiments that followed his remarkable observational studies. He carried these out more than a half century ago among orphanage children prone to pellagra and its then not infrequently fatal consequences. (An irony is that this work of a highly principled investigator today, in my view, would not meet the standards of an Institutional Review Board (144).)

Yet, epidemiologists have remained cautious about experimental intervention. Perhaps they have been deterred on debatable ethical grounds, or perhaps they have been reluctant to sacrifice generalizability and representativeness for specificity and internal validity. After all, the differences in the strength of the inferences about causality drawn from different types of design are not qualitative, but a matter of degree. Aside from the ethical questions always present, a number of other considerations weigh. In time, effort, and expense, the scale of a preventive trial must be thought of in much the same terms as a longitudinal study. As with cohort studies, these are best justified when there is a closely specified hypothesis, and when the same answer cannot be got by other means. It goes without saying that both the intervention and the outcome must be well defined and measurable.

The decision about whether to run a prophylactic trial may be far from self-evident. A New York trial of prenatal nutritional supplementation to prevent low birth weight (145) had a precise hypothesis and a well specified intervention and outcome. It might have been argued—and some reviewers of the initial proposal did argue—that the answer could be obtained through observational studies (144). In truth, reductions in birth weight following prenatal exposure to famine had been demonstrated in at least three studies of World War II experience (93). Nonetheless, when it came to doing the New York study (145), two forms of prenatal supplementation failed to raise birth weight significantly in women at high risk of low birth weight. In fact, the high protein supplement produced an excess of prematurity, of low birth weight, and of newborn deaths.

A lesson here is that the success of an intervention aimed at counteracting the observed effects of a deficiency, or some other adverse factor, can never be taken for granted. For one thing, many variables interact to cause an outcome. Interventions select but one or a few of these variables, thereby ignoring or bypassing unknown interaction effects. Even if the selected experimental interventions are exactly the obverse or counterpart of the adverse factors—and a degree of speculation must reside in the judgment that they are—the effect on the given outcome is of necessity uncertain. Results that always confirm existing hypotheses of course preclude science, which is sustained only by its subversive challenge to existing beliefs. As seen in this study of prenatal nutrition (145), as well as in the University Group Diabetes
Program (138), this uncertainty always includes the possibility of no effect or, more disconcertingly, of an unanticipated harmful effect. Wisdom is to expect the unexpected.

With prophylactic trials aimed at preventing chronic disease, sample size tends to be a limiting factor. Vast numbers may be needed to confer the statistical power to demonstrate significant effects. As a result, the sponsors of recent preventive trials have tended to be government agencies, since risky and expensive multicenter projects are unlikely to be afforded in any other way. Thus, over the past decade, the National Heart, Lung, and Blood Institute has launched several bold experiments in prevention. Prominent among these are the Hypertension Detection and Follow-up Program (146-149) and the Multiple Risk Factor Intervention Trial (150).

The Hypertension and Detection Follow-up Program screened defined populations at 14 centers \((n = 178,009)\) to detect 10,940 persons with diastolic blood pressure greater than 90 mmHg, who would be the subjects of a trial. Participants were assigned randomly either to special care, so-called “stepped care,” or to the regular care of their community physician, so-called “referred care.” (About one fourth in both groups were already receiving antihypertensive drugs.) The stepped care group received comprehensive care with emphasis on a hierarchy of drug treatments.

Five-year mortality was first reported as an end point (146). The results show a significant reduction in mortality in the experimental stepped care group. Most notable, both for its magnitude and for the size of the group, is the effect in those with mild hypertension (diastolic blood pressure 90–104 mmHg).

It is premature to make a final judgment from the findings for this single end point, but one may raise a number of questions. First, was the result solely the effect of the higher frequency of drug treatment in the experimental group (the explanation preferred by investigators) or was it an effect of the comprehensive care that the group received? In favor of a drug effect is the congruence, with the treatment effect, of the differences in the proportion of participants in the experimental and comparison groups who were using antihypertensive drugs.

Second, were the investigators correct to abjure a placebo treatment group? This decision rested on the ethical ground that antihypertensive drug treatment had previously been shown to reduce the morbidity of males with severe hypertension. Were these grounds legitimate for mild hypertension, or even for women? For these classes, no effects of antihypertensive treatment had been previously demonstrated. Whether women could have been considered a potentially separate biologic class was perhaps as much a political as a moral or scientific question.

Third, one may ask what the implications of the study are for preventive intervention. The sponsors concluded, from the favorable effect on mortality among those with mild hypertension, that everyone with any degree of hypertension should be vigorously treated.

Studies of large questions with large consequences plunge epidemiology into the arena of major policy. On some estimates, treatment of mild hypertension could involve up to 15 per cent of the total population and 30 per cent of adults, or more than 30 million people. On more stringent criteria, not less than 6 or 7 per cent, say 9 million people, would be involved. Indeed, the National High Blood Pressure Education Program, under the auspices of the National Institutes of Health, claimed that as many as 60 million Americans, half of the adult population, have high blood pressure and are candidates for treatment (151). If this prescription were to be followed, one is perhaps obliged to ask what is the proper recommendation for several millions at risk of treatment who belong to the subgroups...
in whom no essential differences in mortality emerged.

The question was brought to attention by the investigators themselves. Their analysis of groups subdivided by race, sex, and age, the subject of a second report, showed an effect on blood pressure level in all groups, but showed no effect on mortality in white women or in persons of both sexes aged 30 to 49 years (147). The results for these two groups cannot be played down for lack of power, as implied by the investigators, which is better than 90 per cent for detecting a reduction of 20 per cent mortality.

Subgroup analysis provides a shaky platform for statistical inference unless the strata analyzed have been defined as conditions for experimental testing a priori. Post hoc manipulation of the data ("data dredging"), while valuable in itself, raises three particular problems: if voluntary behavior such as adherence to treatment is involved, the strengths of nonselective assignment to comparison groups peculiar to an experimental design are abandoned; if an hypothesis was not set up for testing, the analysis departs from the model of an experimental test and becomes exploratory; if different segments of the data are tested, each "look" at a subgroup carries its own probability of a result being observed by chance and alters the subsequent conditional probabilities that attach to the hypotheses under test.

The third data-dredging problem can be rephrased in terms of the search for statistical interaction (which is present when results vary across subgroups). To test systematically for interaction is to reduce greatly statistical power and to raise the likelihood of both false negative and false positive results. In tests for interaction, even random misclassification can produce spurious associations (152).

It turns out that in the Hypertension Detection and Follow-up Program (146), subjects were randomized within three predetermined strata of blood pressure level, but within no other strata. By the strict conventions of hypothesis testing, therefore, the absence of effects in the two special sex and age subgroups is without force upon the general result. This is in contrast with the effect in mild hypertension, a subgroup set up a priori.

Yet, in this instance, the scale of the public health problem lends merit to the plea that subgroups should not be entirely ignored in reaching a conclusion; statistical inference should not stand in the way of logical inference. In publishing results for subgroups constructed post hoc, the investigators were surely moved both by the momentous issues they faced and the potential importance of the variations in the data they had collected.

In order to make a judgment, many epidemiologists will want data from other sources to be brought to bear, as well as more results from the study itself. The effects of stepped care on mortality from mild hypertension withstand additional probing for potential confounding of the results by treatment and by the extent of organ damage prevailing at entry into the trial (148). A major Australian trial (153) of drugs versus placebo for hypertension, randomly stratified by age and sex, also supports the experimental results of the Hypertension Detection and Follow-up Program on mild hypertension (>94 mmHg). The experimental subgroup analysis was ambiguous, however; again, no significant effect was obtained either in women or in the age group under 50 years, although the direction of the result in both these subgroups conformed with a treatment effect.

The supportive conclusions of another analysis of the Hypertension Detection and Follow-up Program—that more than half of the excess risk in the referred care group could be attributed to treatment—remain suspect because of self-selection for compliance with treatment (154). More convincingly, the Program finally demonstrated an effect on a different end point—cerebrovascular disease—that was consis-
tent with average blood pressure levels and held across all age and sex groups (149).

Growing certitude was undone by the Multiple Risk Factor Intervention Trial (150) since it failed to demonstrate an effect on mortality from antihypertensive treatment combined with other preventive measures. The Multiple Risk Factor Intervention Trial aimed to use the high-risk selection strategy to reduce mortality from coronary heart disease by interventions that should reduce multiple risk factors. A trial involving diet alone had been given up as infeasible, mainly because of huge numbers and cost in the face of low rates of compliance and the expense of providing the diet at special food centers (155).

About 350,000 men at 22 centers across the country were screened in order to recruit close to 13,000 participants aged 35–57 years (150). These men, selected for high “Framingham coronary risk scores,” were assigned randomly to intervention. The intervention program consisted of advice about reducing dietary cholesterol, counseling sessions on smoking, and “stepped” drug treatment for hypertension.

Sample size for the study was estimated from realistic anticipation of the reduction that could be achieved in each risk factor and from the supposition that in turn these reductions might reduce coronary mortality by about one fourth. The sample could be relatively small because sample size did not allow for hypotheses about subgroups and because a one-tailed test of significance was used.

The first set of results (150), after an average follow-up of seven years, showed no advantage for the experimental group in total mortality (actually +2 per cent) and a minor nonsignificant advantage in coronary disease mortality (actually −7 per cent). Several interpretations of the results have been considered:

1. The regimen for intervention was ineffectual. If so, it was not ineffectual because of any failure to reduce risk factor levels; better than 80 per cent of the goal for risk factor reduction was achieved. Furthermore, unlike the Hypertension Detection and Follow-up Program, this regimen did not have to compete with the active referral of the control group to treatment elsewhere. Since many high-risk individuals at the ages recruited for the trial are likely to have had advanced atherosclerosis, however, it is quite possible that the attempt to intervene came too late for primary prevention.

2. There were beneficial effects in most groups, but these were counterbalanced by the adverse effects of antihypertensive drugs in a minority of subgroups (a view favored by the investigators). Thus, men in the experimental group and those who had hypertension with electrocardiographic abnormalities actually experienced excess coronary heart disease mortality (+65 per cent). Also, those who had mild hypertension did no better than controls. Once again, however, the analysis of subgroups not created a priori does not by itself controvert the statistical inference of a null result of a given experiment as designed.

3. The regimen was effective, but “contamination” of the control groups reduced statistical power to the point at which the effect could not be detected. The regimen for the experimental group had to compete with sensitization to risk of the control group. Those assigned as controls and their doctors were of course informed about their high risk and followed annually, and their voluntary response was marked. Thus, with regard to risk factors, after six years in the study, blood pressure and cholesterol levels of the control group were reduced by margins almost as large as those of the experimental group. Only the reduction in cigarette smoking distinctly favored the experimental group.

With regard to outcome, the experiment also had to compete with history. Over the period of observation, mortality from coronary heart disease in the United States had declined almost as much as was hoped for from the intervention program. At the
six-year point, the number of deaths in the control group was a little less than half the expectation (219 vs. 442). The reduction in control group risk factors and expected deaths together reduced the chance of detecting an effect on mortality from around 90 per cent to 60 per cent. That is, actual statistical power was low.

Thus, these major and courageous experiments are beset by ambiguity and conflict. Epidemiologists who undertake large-scale tests of hypotheses dear to many, and who produce negative or adverse results, must be ready to face criticism from both friends and enemies. Such issues are likely to be buffeted about by ideology, by commercial interests, or by the constraints of national budgets that ruling economists no longer find tolerable.

To conduct epidemiologic experiments is still to pioneer. Avoidable errors as well as unavoidable inconsistencies are costly but inevitable. At this time, no better means offer for answering sharply posed questions about interventions in a definitive way. Other approaches might be possible under special conditions. For instance, were there to be an adequate data base covering total health care for a number of populations, of the kind that is available for Olmsted County through the Mayo Clinic, quasi-experimental studies could yield valuable answers. Again, case-control studies of exposure to intervention could be mounted around selected end points (129–131). Yet, the effort and expense of large-scale experimental studies cannot be entirely avoided if we are to learn how to do better in sharpening the answers to our questions.

**EPILOGUE**

How then should one assess this period of the history of epidemiology in the United States? The diffusion of new epidemiologic knowledge has contributed to major changes in health behavior and mortality: changes in smoking, in diet, in physical exercise, and in the regulation of environmental and occupational hazards, drugs, and dangerous substances. Yet the developmental history I have described has taken place in the shadow of a paradox.

While epidemiology has flowered, the unprecedented frightfulness of nuclear destruction has beset us. That dread is common to everyone living in our age. But the wonders and terrors of science are not unconnected. Even in its most benign aspect, epidemiology must endure the irony of its ties with the malignant outgrowths of scientific creation. Today’s epidemiologists, like all American scientists, are the affines of the Manhattan Project and Los Alamos.

From those undertakings stems the era of big science, a science wrapped in high technology that can only be sustained by government. Epidemiology shares in this evolution. Much of modern epidemiology must be on a large scale because of the demands inherent in its new ambitions. Like other sciences and despite occasional buffetings, for the most part epidemiology has been cossetted by the federal government.

The sins of science arise from the power of its applications. The discoveries of scientists have consequences, unanticipated as well as anticipated, harmful as well as helpful. As the handmaiden of public health and medicine, epidemiology becomes by definition applied. Thus, epidemiologists have sought to go beyond understanding disease, in order to prevent or control it. In the contemporary United States, the discoveries—or more commonly the ambiguous findings—of epidemiology are seen to have consequences as never before.

To demonstrate a carcinogen or a mutagen is at once to bring regulatory agencies and affected industries into conflict or collusion. The failure to demonstrate a favorable effect, as in the case of nutritional supplementation say, may be the occasion for budget cutting on the part of government, and for hunger on the part of the poor. The failure to demonstrate toxic effects of waste chemicals may form the grounds for relaxed regulation and more
environmental contamination. In studies of air pollution, the tough-minded criticism of serious design faults that detract from reported effects (156) lends political support for reducing standards of air quality; to demonstrate that the critique neglects many considerations (157), such as the ambiguous results produced by weak statistical power and weakly measured variables, may not counter the damage. The probably inflated estimate that 40 per cent of cancers have an occupational source leads to one kind of government action and industry reaction; the possible underestimate that such cancers comprise no more than 5 per cent leads to another kind, or more likely, to none (158, 159).

In such matters in the United States today, epidemiologists testify before congressional committees, official commissions, and judges. Not uncommonly they are to be found, sometimes as paid experts, on both sides of the question. Contention on all these issues has strengthened the rigor of epidemiologic research and inference. The inconsistency of results across studies characteristic of epidemiology raises the pitch of argument. Inconsistency is a fact of epidemiology that must be lived with. If not by chance, it may arise from weaknesses and differences in method, from subtle elaborations of analysis, or from true differences and changes in populations, environments, and agents. Thus, the testing of results by replication under the same and under different conditions is a necessary resort, however tedious.

To a large extent, the federal government’s agencies have become the arbiters of what the important problems are that need research and should be funded. Thus, these agencies mould the direction of the greater part of the national research effort. The incentive for epidemiologists to be responsive is enormous if not irresistible.

Some speculate that the combination of generous funding with guidance in the direction of research dampens imagination. Indeed, some applaud the precarious situation of many researchers today—in which they may have little assurance of continuing funds for themselves or their coworkers—as promoting the survival of the fit.

Yet, one may argue the converse. To scurry for funds certainly requires energy and activity. It is a question whether, in the larger scientific context, that activity is well directed. The once well funded system for training and research of past years has been the key to creating the gratifying number of competent epidemiologists in the country today. There is no reason to think that comfortably funded researchers who have the gift of scientific imagination do not exercise it well.

Greater dangers may arise as corporate sponsors enter the field in response to the federal regulation of occupational health and safety. These sponsors are helping to create a new specialization in occupational epidemiology (alongside the few other specializations such as social, environmental, and genetic epidemiology that take origin from the independent variable). They are bound, by the nature of their economic, organizational, and bureaucratic needs, to prove less disinterested than the National Institutes of Health.

Epidemiologic research takes place within the larger society. Its guiding concepts cannot be disconnected from that context, although they may transcend it. Within those limits, epidemiologists have forged research instruments with ever sharper cutting edges. With an appearance of bland neutrality, however, technique may veil the silent intrusion of values and the shaping of judgments by unrecognized societal forces.

Epidemiology is now a mature academic discipline, and epidemiologists properly espouse the values of science and the search for truth or, at the least, explanation. Those values are not, as many scientists have fondly believed, incorruptible. Our obligation is to cultivate awareness and to see that what is learned is not misused. At the
same time, epidemiologists do well to look to the origins of epidemiology as an applied discipline to sustain another set of values. The founders were zealous for the public weal. If we go only as far as the founders, we shall use our discipline to the best advantage in the prevention of disease and the preservation of health.

References

52. Hutchison GB, Shapiro S. Lead time gained by diagnostic screening for breast cancer. JNCI 1968;41:665-81.
76. Beebe GW. Lung cancer in World War I veterans: possible relation to mustard gas injury and
Epidemiology in the United States


94. Berkson J. Limitations of the application of four-


