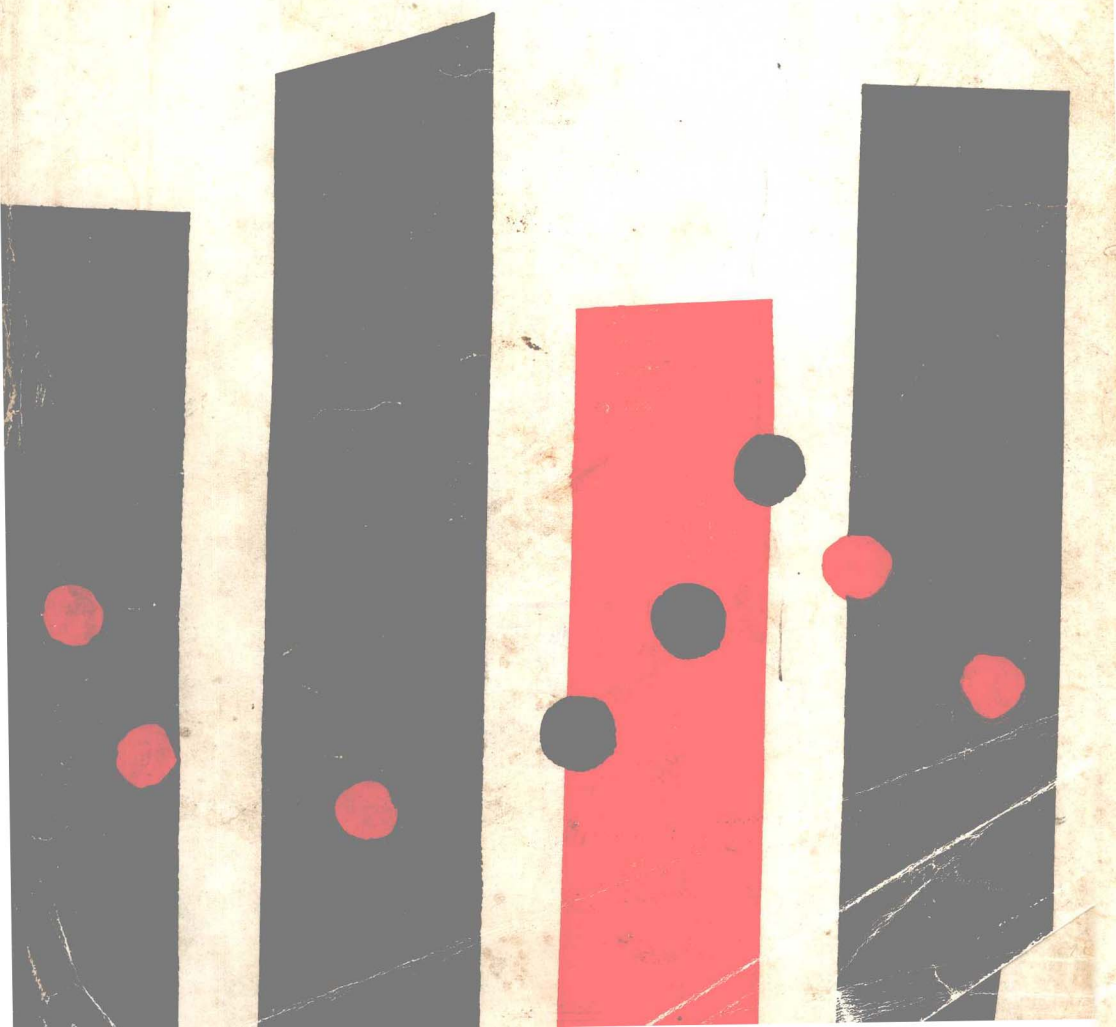
epidemiologic methods

MacMahon/Pugh/Ipsen



Epidemiologic Methods

Brian MacMahon
M.D., PH.D., D.P.H.
Thomas F. Pugh
M.D., M.P.H.
Johannes Ipsen
C.M., DR. MED., M.P.H.

From the
Department of Epidemiology
Harvard University, School of Public Health

LONDON
J. & A. CHURCHILL LTD.
104 GLOUCESTER PLACE, W. 1

LIBRARY OF CONGRESS CATALOG CARD NO. 60–12344

Third Printing

Published in Great Britain by J. & A. Churchill, Ltd., London

PRINTED IN THE UNITED STATES OF AMERICA

Epidemiologic Methods

To

John Everett Gordon

Preface

Epidemiology is widely recognized as the basic science of preventive medicine. This recognition and the greatly increased public support of medical research have found expression in this country in a striking increase in the number of research studies and programs of an epidemiologic type. Naturally, this increase has been most marked for those diseases over which greatest concern is now being felt — arteriosclerotic heart disease, cancer and the mental disorders. While there has been repeated expression of faith in the applicability of the principles of epidemiologic investigation to noninfectious diseases, there is no doubt that the transference of epidemiologic concern from the infectious to the noninfectious diseases has necessitated appreciable modification of methods, if not of principles. We hope, therefore, that this attempt to summarize concepts that seem to us particularly appropriate to the present time will be of service.

Concern with the present by no means implies disregard for infectious disease. Outstanding problems remain unsolved in this field. In addition, our frequent use in this volume of illustrations from studies of infectious disease is testimony to the fact that the present status of epidemiology as a discipline must be credited in major part to its achievements in this area.

In spite of the increasing use of biostatistical techniques in epidemiologic studies, we have avoided, so far as possible, the duplication of material that is readily found in textbooks of biostatistics. We believe that the understanding of the principles discussed here does not require statistical knowledge. However, the reader intending to put some of these principles into practice would do well to sentence himself to a term with a book of elementary statistics, the sentence preferably to run concurrently with the present one.

Whoever wishes to investigate medicine properly should proceed thus: in the first place to consider the seasons of the year, and what effects each of them produces. Then the winds, the hot and the cold, especially such as are common to all countries, and then such as are peculiar to each locality. In the same manner, when one comes into a city to which he is a stranger, he should consider its situation, how it lies as to the winds and the rising of the sun; for its influence is not the same whether it lies to the north or the south, to the rising or to the setting sun. One should consider most attentively the waters which the inhabitants use, whether they be marshy and soft, or hard and running from elevated and rocky situations, and then if saltish and unfit for cooking; and the ground, whether it be naked and deficient in water, or wooded and well watered, and whether it lies in a hollow, confined situation, or is elevated and cold; and the mode in which the inhabitants live, and what are their pursuits, whether they are fond of drinking and eating to excess, and given to indolence, or are fond of exercise and labor.

HIPPOCRATES
On Airs, Waters and Places

Contents

Epidemiologic Methods

1

The Place of Epidemiology in the Investigation of Disease

DEFINITION

Epidemiology is the study of the distribution and determinants of disease prevalence in man.

Two main areas are indicated in the definition. These are the study of the *distribution* of disease (descriptive epidemiology) and the search for the *determinants* of the noted distribution (analytic epidemiology). The first, describing the distribution of health status in terms of age, sex, race, geography, etc., might be considered an extension of the discipline of demography to health and disease. The second, involving interpretation of the distribution in terms of possible causal factors, is the special contribution of epidemiology. Definitions of epidemiology that consider only the descriptive phase do not indicate the unique component of the discipline.

Epidemiology, like most other sciences, has developed from the study of the exotic and the unusual into an inquiry as to general principles, including explanation of the unusual. Epidemiology is now no more restricted to the study of striking outbreaks of disease than meteorology is restricted to the study of hurricanes or astronomy to eclipses of the sun. Yet a modern epidemiologist might still consider his vocation to be primarily a "medical science treating of epidemics" (Webster) in view of two recent trends in the practice of the discipline. These are an expansion

of the concept of what constitutes an epidemic, and a realization that study designed to interpret epidemic patterns cannot be restricted to periods of epidemic prevalence.

The Concept of an Epidemic

The term "epidemic" has until quite recently been almost entirely restricted to the description of acute outbreaks of infectious diseases. However, an acceptable definition of an epidemic might be "the occurrence in a community . . . of a group of illnesses of similar nature, clearly in excess of normal expectation" (American Public Health Association, 1960). This definition encompasses phenomena ranging from the classical pestilences to episodes in the behavior of many noninfectious diseases. The United States, for example, is at the present time in the grip of epidemics of at least two diseases — coronary heart disease and lung cancer — which satisfy as well as any infectious disease the criterion of being clearly in excess of normal expectation. Lung cancer is many times more common in this country than it was twenty years ago, and, although it is now the most common form of cancer in American males, there are many countries where it remains rare.

Epidemic and Nonepidemic Prevalence

A second trend has been growing realization of the value of knowledge of disease frequency and distribution during both epidemic and nonepidemic times, even when the predominant concern may be with the explanation of epidemics.

In the first place, without knowledge of nonepidemic prevalence, how can we know that an epidemic exists? How can we determine whether the frequency of a disease present in a particular population at a particular time is in excess of normal expectancy? Clearly, in order to demonstrate that a particular frequency of morbidity or mortality is in excess of the normal expectancy, it is necessary to know the frequency of the same condition in other populations and in the same population at other times.

In certain instances the existence of an epidemic is perfectly obvious. This is so when the epidemic involves a large proportion of persons, produces evidence of distinctive illness, and occurs

over a short period of time. Thus there is little difficulty in detecting epidemics of cholera, plague, smallpox or the common infections of childhood. In all these cases the disease is familiar, the difference between epidemic and nonepidemic prevalence is large, and the transition is rapid. In contrast, the current risk to an American male of dying of coronary heart disease is quite as large as the risk of death experienced during some of the major epidemics of infectious disease, but the general population remains almost unaware of the existence of an epidemic of coronary heart disease. The slow growth of the epidemic of this disease has served to conceal its enormity from the population.

Even acute epidemics may pass unnoticed if they appear in unfamiliar form. Thus, during the intense London fog of 1952 there was very limited realization of the effects of the fog on the population's health. The full effects were only appreciated when deaths for the period were counted and compared with deaths during the preceding and subsequent periods of the same year and during similar periods of previous years (Figure 1). It then became apparent that the fog was responsible for some 4000 deaths.

Second, finding of an unusually *low* frequency in a particular population at a particular time may be just as significant in understanding the causes of epidemics as the epidemics themselves. For example, the very low attack rates from cholera observed by John Snow (1855) among two groups of people in the center of an otherwise epidemic area led to a strengthening of the belief that the water supply was responsible for the epidemic. Similarly, knowledge of the virtual absence of cancer of the uterine cervix in nuns is of value in the formulation of hypotheses regarding the etiology of this disease.

Finally, in the chronic diseases, which have prolonged upswings and downswings of the epidemic wave, it may be difficult to say whether or not a given frequency qualifies as "epidemic" even if all the necessary comparative information is available. It is common to find a gradient in the relative frequencies of a disease in a number of populations. While the disease may be considered definitely epidemic in those populations at the top of the gradient by comparison with those at the bottom, the status of those in the center may be equivocal. Under such circumstances attempts to correlate quantitative statements of disease frequency

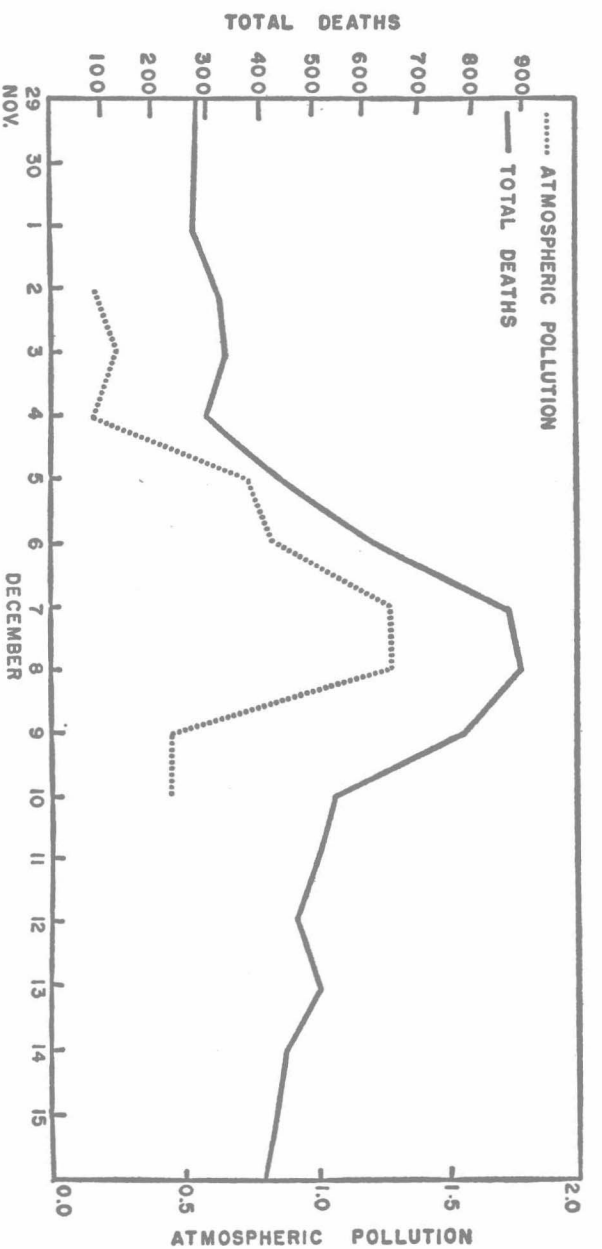


FIGURE 1. Daily numbers of deaths and degree of atmospheric pollution, London, November 29 to December 16, 1952. Atmospheric pollution is measured as parts per million of sulphur dioxide. (Data from the Ministry of Health, 1954.)

with quantitative statements of the frequency of the suspected factors are more precise than attempts to correlate the crude dichotomy of epidemicity and nonepidemicity with the crude dichotomy of presence or absence of specific factors. The former procedure has been followed, for example, in relating characteristics of the diet to rates of coronary heart disease in different populations.

These considerations have led to a great many disease frequency studies which, while having as an ultimate purpose the explanation of unusual prevalences of disease, are not oriented directly toward the study of situations known to be epidemic.

PURPOSES

Knowledge of the distribution of disease may be utilized in one of three general areas — the elucidation of causal mechanisms, the explanation of local characteristics of disease occurrence, and administrative guidance in the provision of health services.

Understanding the Causation of Disease

The most challenging purpose of epidemiology is that of identifying those components of causal mechanisms that enable the formulation of effective preventive measures. This aim encompasses a number of subsidiary, more specific purposes:

(1) Formulation, selection, or rejection of hypotheses that explain disease distribution in terms of specific characteristics or experiences of affected persons.

(2) Testing hypotheses of disease origin through special surveys or other observational studies.

(3) Testing the validity of the rationale on which control programs are based, by the use of epidemiologic data collected in conjunction with the programs. A control program may be considered as an experimental design, the subject of the experiment being the knowledge of causal mechanisms which gave rise to the program.

(4) Distinguishing "entities" of disease and disability through knowledge of disease distribution and of causal factors. On occasion, similarity of epidemiologic behavior may point to etiologic similarity between clinically distinct entities. Con-

versely, difference in epidemiologic distribution for subgroups of an apparent clinical entity may suggest that such subgroups should be regarded as separate disease entities because of probable etiologic differences.

Most frequently, epidemiologists are concerned with these aims as applied to the elucidation of the causation of disease. However, processes of human biology other than those with an end product of manifest disease have been studied. These have included prenatal and postnatal growth, multiple pregnancy, sex determination, intelligence and fertility.

It has been suggested that epidemiology should be concerned as much with those positive components of health implicit in the definition of the word as used by the World Health Organization* as with those negative characteristics that are recognized as frank disease. However, the number of widespread and serious diseases having defined clinical and pathologic characteristics but an unknown etiology appears adequate to occupy the world's epidemiologists for some years to come. Concentration of effort on these latter problems is indicated both because of the great need for knowledge leading to the prevention of these diseases and because of the practical difficulties in the statistical or epidemiologic treatment of concepts that have been neither clinically nor pathologically defined.

Understanding Local Disease Patterns

The preceding purpose dealt with the acquisition of new knowledge about the origin and nature of a disease. In contradistinction, the practicing epidemiologist is often concerned with specific local epidemics or other variations in disease frequency. Here he utilizes what is already known about the etiology of, say, typhoid fever, to explain and deal with an outbreak and to formulate preventive measures suitable to the particular community.

Of course a fine distinction between these two approaches cannot be drawn, since new knowledge may be derived during the course of routine investigation. Nevertheless, in a great deal of "practical" epidemiology, new knowledge of disease causation is not purposefully sought and its advent would be unex-

* "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (W.H.O., 1948).

pected. Typical of this practical epidemiology are the large numbers of investigations of localized outbreaks of food poisoning and salmonellosis undertaken by departments of health with a view to finding the local source of infection, halting the epidemic, and preventing its recurrence. At the present time, because of inadequacies of knowledge concerning other illnesses, such practical uses of epidemiology are largely confined to infectious diseases and certain industrial intoxications.

Administrative Purposes

Knowledge of the frequency of disease in a population serves a number of administrative purposes. For example, it is essential to the logical planning of facilities for medical care. Estimation of the number of hospital beds required for patients with specific diseases (tuberculosis, cancer) or for given segments of the population (infants, the aged) requires knowledge of the frequency and natural history of either those specific diseases or all diseases in those specific parts of the population. In addition, the planning of a variety of studies related to health — clinical, therapeutic, and preventive — requires, for efficiency of design, knowledge of how many cases of a particular disease are likely to be found in a given population during a given period.

Knowledge of the relative frequencies of a disease among subgroups of the population is also useful, since it enables programs and studies to be directed toward the population group manifesting the greatest concentration of the disease. For example, if facilities are limited, it assists in deciding which age, occupational, sex, geographic or ethnic group should be the target for a tuberculosis, diabetes or cancer screening program. Similar considerations apply to the choice of a subgroup of the population for any study, whatever its purpose, that requires the maximum yield of cases for a given size of population studied.

RELATIONSHIP TO OTHER DISCIPLINES

Two characteristics of epidemiology allow us to place this discipline in a logical relationship to the many other disciplines concerned with understanding the causes of disease. These are that epidemiology is primarily an applied discipline, and that its