Epidemiologists can no longer remain satisfied with the specific etiological model which served so well in the study of infectious diseases. A new or modified model of disease causation is needed to deal with current problems, and attention is directed to social science theory as a source for such an endeavor.

SOCIAL SCIENCE THEORY AS A SOURCE OF HYPOTHESES IN EPIDEMIOLOGICAL RESEARCH

John Cassel, M.D., M.P.H., F.A.P.H.A.

SINCE the days of Hippocrates the objective of epidemiological investigation, stated in the most general terms, has been to determine those aspects of man's relationship to his environment that influence his state of health. The crucial question posed by such a formulation lies in the process by which factors in the environment are assessed as being relevant to health states and therefore selected for study. Understanding of the rationale for selection of certain factors as being worthy of study and exclusion of others, equally observable or measurable, can only be gained through a knowledge of the conceptual frame or general theory within which the investigators operated. Thus the concern of the Hippocratic school with factors in the physical environment such as the strength of the prevailing winds, the humidity, temperature, and height above sea level, for example, as explanations for disease occurrence, appears completely logical in view of their theories concerning disease causation. Stated in somewhat oversimplified terms, Hippocrates considered disease to be due to a disturbance in the balance of the four humors in the body—black bile, yellow bile, blood, and phlegm. This balance was determined by the distribution of various attributes of living matter which were categorized as being wet or dry, hot or cold. The distribution of these categories was in turn a function of the relative proportion of the four elements in the environment—earth, water, fire, and air. Thus according to this model, explanation for differential occurrence of various diseases in different populations should be sought in differing proportions of these elements in the environment as indicated by strength of winds, degree of humidity, and so forth.

Viewing the process of selection of variables in this light leads to a further question. What accounts for changes in theory and therefore changes in the factors considered relevant for study? I think it should be emphasized that from a strictly scientific point of view the criterion by which theories are assessed is not one of truth or falsehood, although these terms are frequently applied to theories. Rather the criterion has been (and should be) one of utility, as evidenced by whether the theory has led to findings upon which either intervention is possible or further leads for investigation indicated. In general,

theories have tended to be discarded or modified when the findings predicated by them have failed to provide satisfactory explanations for observable phenomena and thus have provided no leads for intervention or further study.

Germ Theory of Disease

This point can be illustrated by reviewing the evolution of the germ theory of disease, which had very far-reaching consequences. In addition to leading to marked improvements in our ability to intervene, especially as far as treatment of disease is concerned, it led to a new definition of the factors in the environment regarded as relevant, including the discovery of the role of animal and insect vectors in the causation and transmission of disease, and thus has vastly increased our understanding of disease process.

The excitement generated by these discoveries had some unfortunate by-products as well. The new theory was so powerful that for several decades it was believed that microorganisms themselves or at least the general model of causation, of which microorganisms were a specific instance, would be adequate to explain all that was needed to be known about all diseases. This general model of causation held that each disease was caused by a specific agent and that each agent caused a specific disease. As a consequence, attention was focused almost exclusively on increasing our knowledge of the attributes of these specific agents as explanations for all disease processes. Gradually, however, the realization spread that this approach was only providing a partial knowledge of the cause of disease and, for some purposes, not even the most important parts of this knowledge. Increasing information concerning the characteristics of microorganisms has failed to explain why, for example, cholera becomes transformed periodically from a minor pestilence of some oriental bazaar to a raging epidemic; why influenza has periodic episodes of pandemicity; why it is impossible to produce cholera in healthy human volunteers by feeding them the cholera vibrio; or why in some countries only three people out of 100 exposed to the tubercle bacillus develop the disease tuberculosis, while in other countries 15 or 20 or even more out of each 100 exposed will be affected. Not only are answers to such questions unlikely to come from further studies of microorganisms alone, but this model of causation gives very few useful leads as to what other factors need to be studied. Furthermore, even when the search has not been for a microorganism itself, the general model which predicts that there will be a specific cause for each disease has stultified intelligent investigation in many instances.

The relative lack of utility of this mono-etiological model may be further illustrated by an example quoted by Dubos.[1] This example is cancer of the breast in mice. Apparently under natural conditions breast cancer occurs very rarely in this species. Special strains of mice can be bred, however, in which a high proportion of the offspring develop breast cancer spontaneously. Thus a genetic process is involved in the genesis of the disease. It has also been found that a particular virus is present in the milk of lactating mothers of these specially bred strains. If the offspring of such mothers are removed at birth from their mothers and suckled by mice who do not excrete this virus, no breast cancer develops. Thus even though such baby mice possess the genetic characteristic they do not develop breast cancer in the absence of the virus. Furthermore, if mice born from mothers who do not belong to this special strain are removed from their mothers at birth and suckled by mice who do excrete the virus, they still do not develop breast cancer. Thus

exposure to the virus without the genetic predisposition is also without effect. In addition, not all mice born of and suckled by these specially bred mothers develop breast cancer; only female mice do, male mice being immune. Injection of female sex hormone, estrogen, into the male offspring, however, makes them as susceptible as the females. Thus the presence of the genetic factor plus the virus is without effect in the absence of the appropriate hormone. Finally, if mice in which all three factors are present are placed on a low caloric diet, the subsequent incidence of breast cancer is drastically reduced.

The implications of this example are clear. For greater utility we need to modify the mono-etiological model to one which recognizes that factors which may be causal under certain circumstances may under other circumstances be neutral or perhaps even beneficial. Thus the pattern or configuration of factors becomes the crucial issue.

Multiple Causation

The need for a new or modified model of disease causation has been accentuated by the recent inclusion of a further dimension in our concept of the environment. Developments occurring to a large extent since World War II have indicated the potential relevance of the human as well as the physical environment in determining the patterns of diseases prevalent in different populations. The role of these social and cultural processes as etiological factors has been postulated not only for the mental disorders but for a variety of so-called somatic diseases, including tuberculosis, coronary heart disease, rheumatoid arthritis, and hypertension.[2-6] The model of causation derived from the germ theory unfortunately does not provide any useful leads as to the nature of these processes and thus does not aid in the selection of relevant social or cultural characteristics. We do not lack data relating various social and cultural attributes to different disease states, but these findings are often conflicting and contradictory and the interpretations derived from them confusing and unhelpful. The reason for this is that we are trying to select variables and to interpret findings according to a theoretical model which is no longer useful for explaining the phenomena.

Role of Social Science

It thus appears to me that an urgent task requiring the joint efforts of social and health scientists is to develop a conceptual scheme which, by indicating the social and cultural processes of potential relevance to health, will provide leads as to the characteristics to be selected for study and help interpret associations that are discovered.

In order to develop such a scheme it is necessary to explore some further dimensions of our ideas of "cause" as it relates to disease processes. The need for a multicausal framework as illustrated above has been countered by some investigators who maintain that this need is only a reflection of our lack of specific knowledge and the level of abstraction at which we are working.[7,8] According to these sources, current studies of diseases of "unknown etiology" including most of the chronic diseases and mental disorders, for example, are concerned with identifying the vectors of the specific causal agents, in much the same way as polluted water was considered the cause of cholera prior to the discovery of the cholera vibrio. With increasing knowledge the specific agent within this vector will be discovered and we will then have identified "the cause" of the condition. Furthermore, we should recognize that such a cause may be necessary but not sufficient to produce the disease, which would explain why not everyone exposed to this cause becomes

affected. These authors have indeed suggested that a modified series of Koch's postulates be applied to any factors suspected to be etiological in such disease. Such a formulation based on findings in microbiology overlooks an essential premise implicit in the classification of syndromes for which the microorganismal model was most relevant, the infectious disease.

The rationale for combining a series of signs and symptoms into a specific infectious disease is, in most instances, that such a syndrome is a consequence of some previously identified etiological factor or factors. In other words, the present classification of syndromes into specific infectious diseases is an etiological classification arrived at in many instances after the etiological agent had been postulated or discovered. Any classificatory system of disease which is not based upon an etiological hypothesis but upon anatomic distribution of signs or functional change, for example, would thus be inappropriate as a basis for the search for necessary "causes." Thus we would consider it inappropriate today to search for the necessary cause of pneumonia (an anatomical classification) or fevers (a physiological classification) for example. Instead we have regrouped the signs and symptoms into new classificatory schemes based upon their postulated or observed etiological relationship. Some syndromes have been subdivided as, for example, pneumonia which has been divided into pneumococcal pneumonia, virus pneumonias, lipoid pneumonia, and so forth. Other syndromes have been combined, general paresis, chancre, rash, and gummata all now being called syphilis. Given such a classificatory system the identification of some agent as a necessary if not sufficient cause is not surprising, being a consequence usually of the very definition of the disease. By contrast many of the chronic diseases are classified on the basis of their anatomical distribution (cardiovascular diseases, or peptic ulcer for example) or on the basis of behavioral manifestations (psychoneurosis for example). There is no assurance therefore that the existing classificatory system of many of our "modern" diseases is useful for studies designed to determine factors responsible either for the onset of disorder or for failure to recover.

This caution is particularly germane when considering the role of the sociocultural factors in the onset of diseases. A number of relatively recent studies, for example, have indicated that the social experiences of people who subsequently develop tuberculosis or schizophrenia or who commit suicide are remarkably similar.[2,9,10] Such people share in common the fact that they frequently come from a broken family, that they live in an area in which they are a distinct minority not accepted by the dominant majority, that they have had an excessive number of residential and occupational changes, that they are more likely to be single, divorced, or widowed than is the rest of the population, and that they have been subjected to mounting life stress without any period of remittance. People who develop manic-depressive psychosis on the other hand do not appear to have many of these characteristics. On this basis it would thus appear potentially useful to classify schizophrenia with tuberculosis and suicide as one syndrome rather than regard it as a similar disease to manic-depressive psychosis (i.e., both being classified as a psychosis) and distinct from tuberculosis.

The modification in current epidemiological strategy that these ideas would suggest is that more studies should start with people exposed to a postulated etiologically relevant process and determine the spectrum of disorders that are a consequence rather than attempt to determine the antecedent necessary processes of specific diseases, classified ac-

cording to current usage. The findings of Hinkle and his colleagues in their studies of industrial employees add additional support to the utility of this view.

Categories of Causes

Finally, in considering the concept of "cause" as applied to disease, I think that greater attention must be given to the possibility that those sets of "causes" which are responsible for the onset of conditions may be very different from those responsible for the lack of recovery from those conditions. To the best of my knowledge, this important distinction between the categories of causes was first suggested by Halliday in 1943[11] but has not as yet found general acceptance. Halliday points out that knowledge concerning the first category of causes, those responsible for the onset of conditions is of vital importance in the prevention of new cases occurring. Such knowledge may be irrelevant, however, for the adequate treatment of already manifest cases. Similarly, knowledge concerning those causes responsible for lack of recovery may be essential for therapeutic purposes but of little if any use for the prevention of new cases. This point may be clarified by the use of a simple illustration. Increasing knowledge concerning insulin metabolism has materially improved our ability to treat cases of diabetes. No matter how sophisticated our knowledge of this type of "cause" of diabetes, however, it provides no guidance as to how to prevent the next case of diabetes occurring. If, however, to take a hypothetical example, we had information to the effect that immigrants to a new country developed the diabetes rates of that country to the extent that they accept some of the customs of their host country, and we knew which these customs were, we would be in possession of information which would help prevent diabetes but which might not be very useful for the treatment of diabetics.

To put this in other words, by studying those causes of disease which provide useful knowledge for therapeutic purposes and by expecting that these same factors will be responsible for the onset of conditions, we may be guilty of the logical fallacy of saying that because water quenches fire the cause of fire is therefore lack of water. In developing a model which will lead to the identification of etiologically relevant social and cultural processes, therefore, it is necessary to be explicit as to which category of etiology we are concerned with.

Using these ideas as a general framework, we have attempted to develop a more specific conceptual scheme which would lead to useful epidemiological studies. As a starting point we attempted, by drawing on selected biological, psychological and social theories, to define some of the general social processes that could be regarded as potentially deleterious to health and as of the present we are exploring two such sets of processes. First, we were intrigued by the possibility that disproportionately rapid rates of change in any one of the four linked open systems described by Caudill,[12] the physiological, psychological, social, and cultural, could, by producing strains on the others, lead to breakdown. Accordingly, we have been searching for situations in which there has been rapid change in one of these systems, the social system, but minimum change in the cultural system of the exposed population, and exploring the consequences on the physiological and psychological systems in terms of selected health indexes. The specific instance of rapid social change we have selected has been urbanization and industrialization. We have conducted two studies along these lines which are reported in full elsewhere[13-15] and which will therefore only be briefly summarized. In the one we compared selected aspects of the health of rural mountaineers who were the first of their

family to engage in industrial work with their co-workers drawn from the same ethnic stock and from the same mountain coves doing the same work in the same factory for the same wages, but who were the children of parents who had worked in this factory before them. We postulated that the recent rapid social change that the first group had undergone would increase the likelihood of incongruity between the demands of the social system and the culture they brought with them and predicted that they would thus have poorer health indexes.

In the second study, instead of studying people who had changed, we selected stable groups in which the social situation had changed around them. Accordingly we selected rural residents in North Carolina living in counties with differing degrees of urban growth. Our predictions were that the postulated incongruity would be greatest the larger the size of the city in the county in which these rural residents lived, and consequently there would be higher rates of ill health under these circumstances. In both studies our predictions were supported by the data. No matter how we measured ill health, by absence from work and response to the Cornell Medical Index in the first study, or by death rates including total deaths, deaths from all heart disease and coronary heart disease in the second study, the groups with greater likelihood of experiencing this incongruity had the higher rates.

Further Explorations

We do not of course regard these studies as definitive evidence supporting our scheme but have been sufficiently encouraged to initiate further projects to test it more directly. These have just begun and we have no data as yet.

Finally we are exploring a second process of potential relevance to health, the degree of integration within a social group. Our reasoning here is that the consequences of any deleterious set of circumstances need not be expressed in maladaptation of the physiological or psychological system if there are meaningful groups through which such individuals can derive adequate emotional support. To test the feasibility of this postulation my colleagues, Drs. Norman Miller and H. A. Tyroler, have conducted some very preliminary analyses of work group relationships in an industry. They selected two groups of hourly paid male employees doing similar jobs. The one group (Group A) were those people who in common with the rest of their work group worked either fixed or rotating shifts. Thus they would have a constant set of fellow workers with whom they could interact. The second group (Group B) either worked fixed shifts in work groups predominantly made up of people who changed shifts periodically, or they themselves changed shifts among groups whose shifts were fixed. Thus they were not afforded the opportunity for close continuous interaction with the same group of fellow workers. As of now Drs. Miller and Tyroler have only examined one health index, the level of serum cholesterol in these two groups. Considering the crudeness of the index of availability of group support, the results are rather startling. As compared to the levels of serum cholesterol in the total population of employees, and after adjusting for age differences between the two groups, it was found that the proportion of Group B men who were hypercholesteremic (that is were in the upper highest quartile of values for the plant as a whole) was twice as great as was the proportion of Group A men. Conversely the proportion of Group B men in lowest quartile of serum cholesterol values was only half that of the Group A men.

We would be the first to recognize that these preliminary studies which I

have summarized do not prove in any formal sense the utility of such schemes. They do however provide what we consider to be valuable guidelines for further study by indicating the types of variables and situations we should select and help make sense of some of our findings together with those in the literature.

REFERENCES

1. Dubos, Rene. Mirage of Health. New York, N. Y.: Harper, 1960.
2. Holmes, Thomas H. "Multidiscipline Studies of Tuberculosis" in Personality, Stress and Tuberculosis. Phenias J. Sparer (Ed.). New York, N. Y.: International Universities Press, 1956.
3. Lee, Richard E., and Schneider, Ralph F. Hypertension and Arteriosclerosis in Executive and Non-Executive Personnel. J.A.M.A. 167:1447, 1958.
4. Friedman, Meyer, and Rosenman, Ray H. Changes in Serum Cholesterol and Blood Clotting Time in Men Subjected to Cyclic Variation of Occupational Stress. Circulation 17, 1958.
5. Scotch, Norman A., and Geiger, H. Jack. The Epidemiology of Rheumatoid Arthritis. J. Chronic Dis. (In press.)
6. ————. The Epidemiology of Essential Hypertension. (Manuscript.)
7. Yerushalmy, J., and Palmer, Carrol E. On the Methodology of Investigations of Etiologic Factors in Chronic Diseases. J. Chronic Dis. 10:27, 1959.
8. Lilienfeld, Abraham. On the Methodology of Investigations of Etiologic Factors in Chronic Disease—Some Comments. Ibid. 10:41, 1959.
9. Chapman, Loring F., et al. Human Ecology, Disease and Schizophrenia. Am. J. Psychiat. 117:193, 1960.
10. Hare, E. H. Mental Illness and Social Class in Bristol. Brit. J. Prev. and Social Med. 9:191, 1955.
11. Halliday, James L. Principles of Aetiology. Brit. J. M. Psychol. 19:367, 1943.
12. Caudill, William. Effects of Social and Cultural Systems in Reaction to Stress. Social Science Research Council, Pamphlet No. 14 (June), 1958.
13. Cassel, John; Patrick, Ralph; and Jenkins, David. Epidemiological Analysis of the Health Implications of Culture Change: A Conceptual Model. Ann. New York Acad. Sc. 84:938, 1960.
14. Cassel, John, and Tyroler, Herman A. Epidemiological Studies of Culture Change. I. Health Status and Recency of Industrialization. Arch. Environ. Health 3:25, 1961.
15. Tyroler, Herman A., and Cassel, John. Epidemiological Studies of Culture Change. II. The Influence of Urbanization on Cardiovascular Mortality Rates of Rural Residents. (Manuscript.)

Dr. Cassel is associated with the Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, N. C.

This paper was presented at a conference sponsored by the National Institutes of Mental Health on the Relation of Social Factors to the Etiology of Disease, November 28-December 2, 1962.