

---

## 2. *Causes and Entities of Disease*

*Brian MacMahon and  
Thomas F. Pugh*

BASIC to any useful system of disease prevention or therapy is the construction or definition of disease entities. These entities are created by classifying ill persons according to characteristics thought to have relevance in respect to the cause, mechanism, or course of their illnesses. The members of such a group are said to have a specific disease or disease entity.

The arbitrary features in the practice of creating disease entities are frequently not recognized. Ill persons fall into many distinguishable groups—so many that one could readily construct a classificatory scheme that embraced as many disease entities as ill persons and consequently was useless with respect to generalizations of any sort. The disease entities that are in fact recognized have been selected, from the innumerable possibilities that exist, on the basis of usefulness for prevention or treatment or on the basis of medical tradition.

For purposes of prevention, it is most useful to identify as disease entities those in which causal factors are the definitive characteristics. Again, there are innumerable factors in the environment and in genes that are causative of disease. The disease entities created will then vary considerably according to the causal factors that are selected as the basis of classification. The selection is arbitrary and is determined by the utility of the disease entity created or by the priority of discovery of the association of a particular causal factor with a group of ill persons. Significant new knowledge of etiology may bring about major revisions of the boundaries of existing entities.

It is evident therefore that concepts of what constitutes a cause of disease and how diseases should be classified for preventive purposes are closely intertwined.

## CONCEPTS OF CAUSE

An appropriate definition of cause is that of J. S. Mill [3]: "The cause, then, philosophically speaking, is the sum total of the conditions positive and negative taken together; the whole of the contingencies of every description, which being realized, the consequent invariably follows." Note that Mill's definition is empiric—given certain conditions, other conditions are observed—and does not require knowledge of the mechanism of the relationship. As Hume [2] earlier remarked: "We are never able, in a single instance, to discover any power or necessary connection, any quality which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find that one does actually, in fact, follow the other." In terms of Mill's definition, it is evident that the goal of understanding the cause of any disease or event clearly is not attainable. The only practical aim can be the identification of some of the components in the infinite causal web. While it is hopeless to attempt to answer the question "What is the cause of this disease?", it may be possible to obtain an answer to such questions as "What are some of the causal contingencies in this disease?" or "Is this factor a cause of this disease?" In the last question, the query is whether the factor A is one of the contingencies that in toto constitute Mill's cause of event B. If it is, then A being present and all the other contingencies operative, B will invariably follow.

In practice, it must be determined whether A is one of the causal contingencies of disease B in the absence of knowledge of what the other contingencies are. In the absence of such knowledge, an assumption may be made that the other contingencies are distributed equally between persons who experience A and those who do not. If this is the case, and if A is a cause of B, a higher frequency of B will be found in A persons than in not-A persons, even though B is not found in all A persons. The first requirement in determining that A is a cause of B, then, is that a statistical association exists between A and B.

Not all instances of statistical association are indicative of a causal relationship. That A is a cause of B implies that B would not have occurred in the absence of A, a criterion not met by many statistical associations. For example, both cholera and typhoid fever may depend on poor sewage disposal and thus be statistically associated, but cholera would not be considered a cause of typhoid fever or vice versa. A statistical association can be defined as causal when there is an association between two categories of events such that alteration of one (the cause) is observed to be followed by a change in the frequency or quality of the other (the effect). If the supposed cause cannot be, or has not been, altered, an association may be classed presumptively as causal when it is believed that, had the cause been altered, the effect would have been changed.

## INVESTIGATION FOR CAUSAL RELATIONSHIP

In investigating the possible causal nature of an observed association between two categories of events, an experiment may be designed to determine whether in fact a change in one follows a change in the other. In such an experiment, the subjects are divided at random into two or more groups to ensure that, within the limits of chance variation (which can be estimated), the groups are highly similar. One group is exposed to the suspected factor, and the other is not. Because of the random nature of their selection, it may be assumed that the groups are exposed equally to all other contingencies, and differences in outcome (B) between the groups are judged to be effects of the manipulated factor (A). This model embodies the concept underlying tests applied with increasing frequency to the evaluation of drugs, other

therapeutic procedures, and such preventive measures as vaccines and fluoridation of water supplies.

In many instances, random assignment to experimental groups is impracticable, unethical, or undesirable because of medical tradition. Here judgment whether the association is causal becomes difficult, because it must depend on observational data, and the validity of the assumption regarding random distribution of the "other contingencies" is always open to question. Three types of evidence may be brought to bear on the judgment:

1. Strength of the statistical association. The frequency with which two types of events would be expected to occur together by chance within a specified time or geographic distance can be calculated from their separate frequencies. The more frequently the two are in fact observed together, compared with the expected chance frequency, the more probably their association is explained by a causal relationship between them.

2. Time sequence. Instances of the supposed cause should precede instances of the supposed effect.

3. Consistency with existing knowledge. Existing knowledge often suggests that a mechanism linking the proposed cause and effect is conceivable. The existence of a series of known cause-effect associations (A:B, B:C, C:D) would make a proposal of an overall association (A:D) seem not only understandable, but even probable. Existing knowledge may also suggest analogies with currently accepted causal associations that make the causal quality of the one under consideration more reasonable. Thus, there may be considerable resistance to recognition and acceptance of the first example of a new type of factor as a disease-causing agent, but once the first example has been accepted (for example, a microorganism or a genetic element), other agents of the same type are more readily recognized and accepted.

A strong case as to the causal nature of an association can often be built from observational evidence. In the absence of controlled experiment, however, it is possible always to doubt the causal nature of an association and to offer other explanations of the observations. When controlled experiment is not possible, the demand for absolute proof of causation in diseases of human beings is unrealistic. A critical point of view in interpreting observational relationships as causal is prudent, but at some point in the accumulation of evidence, it becomes wiser to accept the causal hypothesis as a basis for action than to continue to debate its validity. Controversy should be limited to discussion of where this point lies, rather than focus on an unrealistic demand for absolute proof of one hypothesis or the other.

## NOMENCLATURE AND CLASSIFICATION

A nomenclature is a list of names or designations, with specification as to the characteristics of the things subsumed by the individual designations. A classification is a systematic ordering of such a list. Since the processes of constructing nomenclature and classification are not clearly separable, the latter being dependent on the former, the term classification will be used loosely to include both processes in this discussion.

There are two primary axes of classification of ill persons: manifestational and causal. Manifestational criteria group patients having in common one or more specific manifestations of illness: symptoms, signs, or laboratory determinations. Examples of manifestationally defined illnesses include the common cold, gastric ulcer, and carcinoma of the lung. Causal criteria group patients according to some common prior experience, judged to be of a causal nature. Examples are tuberculosis, avitaminosis, and suicide. All diseases include some manifestational element,

if only to the extent that illness is itself a manifestational criterion. Thus, the category of suicide is causally defined, in the sense that it defines a particular cause of death: one's own hand. However, since the state of death is required for inclusion in the category, it is also manifestationally defined. Similarly, the term tuberculosis is usually restricted to persons who, in addition to harboring a causal agent, the tubercle bacillus, also manifest illness.

Originally, disease classification was based primarily on manifestational criteria, and its basis remains predominantly manifestational today. A change to causal criteria is introduced when etiologic factors have been identified as significant and offer promise of major therapeutic or preventive advantage. The recognition of a significant etiologic factor, however, does not necessarily lead to its inclusion in the classification, as it may seem equally useful to continue the manifestational classification. Thus, the identification of cigarette smoking as an important cause of squamous carcinoma of the bronchus has not led to any revision of classification since, for the purpose of therapy—still the dominant purpose of medicine—it is more useful to group patients with bronchial carcinoma together, regardless of cause, than to classify all the different manifestations under "cigarette smoker's disease."

Important to keep in mind is the fact that selection of a particular causal component for the purpose of disease classification depends on its usefulness. The supposition that a chosen component has some more essential relationship to disease than other components may be false, for such a supposition leads occasionally to the misconception that the selected factor is the cause (e.g., that Mycobacterium tuberculosis is the cause of tuberculosis). The importance of this recognizability of introducing new classifications should be judged by their utility compared to alternative classifications and not on an idea that one classification may be more correct or natural than another.

The pedestrian, struck by an automobile and dying with a ruptured spleen shortly after admission to hospital, is an example. The pathologist may ascribe the death to splenic rupture, the internist to shock, the surgeon to delay in diagnosis in the admitting room. The Registrar of Vital Statistics may be content to assign the death to "Motor vehicle accident involving pedestrian." The highway engineer, in defending his next annual budget, may attribute the death to lack of adequate separation of pedestrian and vehicular traffic, while the engineer responsible for automobile design may count the case among those due to brake failure. Others may point out that the death would not have occurred if the victim had been 25 rather than 65 years old or if his reaction time had not been seriously reduced by alcohol. Each professional observer has selected a different contingency from Mill's causal complex and used it as the basis for his classification. In so doing, he has inferred a special causal relationship of one of the contingencies to the death. Each classification is as valid as another because it is useful for the particular purpose the professional worker had in mind. If the same detail of knowledge existed concerning the etiologic factors in coronary heart disease, no doubt a parallel variety of possible classificatory schemes would exist.

Change from a manifestational to a causal axis of classification may result in a major regrouping of impaired individuals, both in the direction of combining groups with diverse manifestations and of dividing an apparently homogeneous manifestational group. Such a regrouping of impaired persons is exemplified in the commonality of interaction with Treponema pallidum. The result is a grouping of one direction by the establishment of syphilis as a disease entity, based on the such diverse manifestations as chancre, some maculopapular lesions of the skin, aortic insufficiency, and progressive dementia. In the other direction, the term

acute yellow atrophy of the liver, no longer in use, once included patients now ascribed to etiologies as separate as toxemia of pregnancy, infectious hepatitis, and various chemical intoxications. Note that the use of the etiologic axis of classification resulted in both instances in entities with diverse manifestations, those of toxemia of pregnancy not being limited to derangement of the liver. Further, no single manifestational entity was reclassified in its entirety to a specific etiologic entity. Thus, not all cases formerly designated as acute yellow atrophy are now classed as infectious hepatitis; nor are all patients with maculopapular rashes, aortic insufficiency, or progressive dementia included with syphilis sufferers.

In this connection, it may be noted that the first of the so-called postulates of Koch<sup>1</sup> is an example of circular reasoning. This postulate states: "The micro-organism should be found in all cases of the disease in question, and its distribution should be in accordance with the lesions observed." The circularity of the reasoning lies in the fact that the postulate refers to "the disease in question" as if it had been thought of as an entity before the entire range of effects from the microorganismal element of cause had been identified. In fact, evidence is lacking that the full range of such entities can be recognized before the establishment of causal connection between a disease agent and ill health.

#### CURRENT CLASSIFICATIONS OF DISEASE ENTITIES

While it is not difficult to understand that the theoretical basis for selection of a classification should be its usefulness, it is very difficult, in selecting a particular classification for general use, to reconcile the different requirements of the many purposes it must serve.

Two disease nomenclatures or classifications are in widespread use in the United States: the Standard Nomenclature [6] published for the American Medical Association and the International Classification of Diseases [8] published by the World Health Organization.

Introduced in 1930 at the initiative of the New York Academy of Medicine, the Standard Nomenclature, now in its fifth edition, is widely used for indexing hospital patients. It "attempts to include every disease which is clinically recognizable." The very exhaustiveness of the nomenclature, however, has led to the situation that a disease of interest to a research worker may be represented by several terms.

The International Classification, in use since 1900, is now in its eighth (1965) revision. Since 1946 the World Health Organization has played a major role in the development and use of this classification, but the successive revisions remain the responsibility of a decennial International Classification Revision Conference representing the user nations. The International Classification is used in the United States and most other countries primarily for the preparation of national statistics on causes of death. Since it is less exhaustive than the Standard Nomenclature and contains many such residual categories as "Injuries of other and unspecified nature," it has been considered unsuitable for clinical use.

An adaptation of the International Classification [4], prepared under the direction of the U.S. National Center for Health Statistics, supplies sufficient specificity for clinical purposes and simultaneously retains the greater statistical simplicity of the parent classification. The adaptation is expected to supersede the Standard Nomenclature in hospital practice.

<sup>1</sup>A series of three criteria used by Koch in demonstrating the pathogenicity of *M. tuberculosis* and referred to as Koch's postulates by later microbiologists. They are itemized by Topley and Wilson [7].

## DIVERSITY OF EFFECTS

Skepticism as to the existence of causal association has been expressed when a proposed cause has been found to be statistically associated with more than one established manifestational entity [5, 9]. For example, the fact that cigarette smoking is associated not only with bronchial cancer but also with chronic bronchitis, coronary heart disease, and even bladder cancer and other diseases has been put forward as an argument that the association with carcinoma of the bronchus is not a causal one, since no cause, it is claimed, should have so many different effects [1]. The assumption that traditional schemes of classifying diseased people, based on manifestational criteria, should be congruent with schemes based on etiologic studies or delays reconceptualizations that may have both preventive and therapeutic usefulness.

No doubt there is a tendency toward clustering of specific clinical features and other manifestations among patients afflicted with a particular cause of disease, but this by no means amounts to complete correspondence of manifestational and causal groupings. Patients with tuberculosis, for example, include those with a wide variety of manifestations: persons with affections of the bones, lungs, meninges, and skin. Many of these persons would never have been thought of as belonging with others who are currently their nosologic companions except for the presence of *M. tuberculosis* in their lesions. At the same time other causes (e.g., Histoplasma capsulatum) produce similar polymorphous manifestations often indistinguishable from those "caused" by *M. tuberculosis*.

One does find diseases in which there is very high association of a particular cause with a particular effect. For example, in certain genetically determined errors of metabolism, a series of mutations of specific single genes is associated with a corresponding series of single enzyme deficiencies that become manifest in quite specific biochemical abnormalities, such as the inability to metabolize phenylalanine. Note that even here the specificity results in part from the arbitrary selection of cause, for the causes responsible for the original genetic mutation in a series of cases of phenylketonuria may be quite varied. Further, a group of patients having an identical enzyme deficiency may show quite varied degrees of mental defect and even include persons with average intelligence. Similarly, there is probably a close, although not complete, correlation between patients defined because of their having a first exposure to measles virus and those defined by having clinical measles, as well as between the members of a group defined because of having been bitten by a rabid animal and those later regarded as the cause of certain factors may suggest that certain factors can be regarded as the necessary cause of particular diseases, with other causal factors relegated to subsidiary or contributory roles. In fact, however, "necessary" causes are necessary only by definition; for example, *M. tuberculosis* is the necessary cause of tuberculosis, as automobiles are of automobile accidents. A specific genetic deficiency may be thought of as the necessary cause of phenylketonuria, but, in a society in which all were homozygous for the phenylketonuria gene and diets varied in their phenylalanine content, it would be more useful to think of the disease as determined environmentally—by phenylalanine in the diet—and to regard this as the essential factor. The necessary cause is the chosen cause.

In contrast to the above examples, the majority of causal agents that are chosen as criteria for constructing disease entities are associated with a great diversity of clinical, pathological, and biochemical patterns. These patterns often have not been considered part of "one" disease until they have been found to be regularly associated with the chosen agent. Causal agents showing wide variation in their effects

from person to person include *M. tuberculosis*, *T. pallidum*, alcohol, poverty, the automobile, and "one's own hand."

Past experience, then, suggests that arrangements of ill persons by their manifestations may identify groups that have at least some degree of homogeneity with respect to causal factors and that at least form a useful basis for investigation of cause. Nevertheless, causal factors of disease, when identified, not uncommonly have effects that cross the boundaries of adjacent manifestational groups. In so doing, they produce new groupings of ill persons, some of which alarm nosologists who use tradition as their guiding principle.

### CAUSATION AND PREVENTION

Not all demonstrated causal associations are useful bases for preventive measures. This limitation is exemplified by the following considerations:

1. The cause must be amenable to manipulation. Thus, the identification of viral agents in human cancer does not offer the prospect of as rapid a development of preventive measures as might be expected from the identification of microbiological agents in other categories of disease, since special problems are involved in the production of vaccines. As another example, practical measures for preventing genetic damage have not yet been developed.

2. Manipulation of the causal factor must be acceptable. Alcohol, the cigarette, and the automobile are each responsible for an enormous burden of mortality and morbidity in present-day society, but society seems unwilling to give up the benefits associated with them. Lack of acceptability may depend on factors quite unrelated to health.

3. Almost any interference with the existing ecological balance will have effects on health other than those intended or desired. While the "great sanitary awakening" at the turn of this century did rid the Western world of cholera, typhoid, and a great many other diseases, it may also have fostered poliomyelitis and possibly other diseases against which the population was formerly protected because of its unsanitary living conditions and consequent early immunization. Measures such as routine chest x-ray seem imperative under certain conditions and contraindicated under others. The extent of the side effects of preventive measures usually is not known at the time they are introduced. Therefore, periodic surveillance and evaluation of the total effects of such measures are required.

Realization that the directness of a causal association is not necessarily related to its utility for prevention is also important. Directness of causal association is a relative term, dependent on the current state of knowledge of the mechanism of a particular association. Swamps are causally associated with malaria. When this association was found to be explainable because mosquitoes breed in swamps, the association of malaria with mosquitoes was then considered the direct one, and the association with swamps indirect. With the discovery of the malaria parasite, the association with the parasite was considered direct and that with mosquitoes indirect. No doubt there are presently unknown components of the parasite that, when determined, will be considered the more direct causes of malaria ("swamp fever") than the whole parasite.

The more direct the association, the more specific the preventive measure is likely to be and the fewer the anticipated side effects. Thus vaccines, directed specifically against certain infectious agents, tend to produce less ecological disturbance than measures based on indirect aspects of the disease mechanism, for example, sanitary measures. Nevertheless, associations that are recognized as

being indirect can form the basis of effective preventive measures, either because other direct associations are not known or because manipulation of the more direct associations is impracticable or inconvenient. The drainage of swamps is still a useful preventive measure against malaria, and the control of cholera in the Western world depends on general sanitary measures rather than immunization. Knowledge of causal associations that do not offer preventive possibilities, either because from a practical point of view the cause is unalterable or because the side effects are unacceptable, is nevertheless important. Such knowledge aids the study of other potentially causal associations, some of which may be alterable. For example, knowledge of genetic determinants of a disease can facilitate identification of environmental factors involved in its occurrence or severity.

## REFERENCES

1. Berkson, J. Smoking and lung cancer: some observations on two recent reports. *J. Amer. Statist. Ass.*, 53:28, 1958.
  2. Hume, D. *Treatise of Human Nature*. (1739.) Selby-Bigge, L. A. (Ed.). Oxford: Clarendon Press, 1896.
  3. Mill, J. S. *A System of Logic, Ratiocinative and Inductive*. (5th Ed.) London: Parter, Son and Bowin, 1862.
  4. National Center for Health Statistics. *International Classification of Diseases, Adapted for Indexing Hospital Records by Diseases and Operations*. Vols. 1 and 2. (Revised Ed.) Pub. Health Service Publ. 719, Washington: Govt. Printing Office, 1962.
  5. Sartwell, P. E. On the methodology of investigations of etiologic factors in chronic diseases. Further comments. *J. Chronic Dis.* 11:61, 1960.
  6. Thompson, E. T., and Hayden, A. C. (Eds.). *Standard Nomenclature of Diseases and Operations*. (5th Ed.) New York: McGraw-Hill (for the American Medical Association), 1961.
  7. Topley, W. W. C., and Wilson, G. S. *The Principles of Bacteriology and Immunity*. (2nd Ed.) Baltimore: William Wood, 1936.
  8. World Health Organization. *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death*. Vols. 1 and 2. 7th (1955) Revision. Geneva: World Health Organization, 1957. (The eighth revision was in process in 1966.)
  9. Yerushalmy, J., and Palmer, C. E. On the methodology of investigations of etiologic factors in chronic diseases. *J. Chronic Dis.* 10:27, 1959.
- Hempel, C. G. Introduction to problems of taxonomy. In Zubin, J. (Ed.), *Field Studies in the Mental Disorders*. New York: Grune & Stratton, 1961.
- Lorr, M., Klett, C. J., and McNair, D. M. *Syndromes of Psychosis*. New York: Macmillan, 1963.
- MacMahon, B., Pugh, T. F., and Ipsen, J. *Epidemiologic Methods*. Boston: Little, Brown, 1960. Chaps. 2 and 3.
- Nowell-Smith, P. H. Causality. In *Encyclopaedia Britannica*. Chicago: Encyclopaedia Britannica, 1965.

## ADDITIONAL READING