

## PROBLEMS OF IDENTIFICATION OF CASES OF CHRONIC DISEASE

P. E. SARTWELL, M.D.

**T**HE problem of diagnosis and classification arises whenever we try to study the epidemiology, or for that matter the clinical characteristics, of any disease, acute or chronic. The statistician and epidemiologist are concerned only with things that are countable. Bradford Hill quotes Bartlett as saying: "In so far as things, persons, are unique or ill-defined, statistics are meaningless . . . our arithmetic is useless unless we are counting the right things."

Recent years have seen a greatly improved appreciation of the inadequacies of diagnostic tests, and consequently of diagnoses, even in situations where it had been confidently assumed that errors were trivial. As an illustration, you are all familiar with the studies of the reproducibility of chest x-ray readings, so important in a screening program. This is a real advance. Much work, however, needs to be done in validating all sorts of commonplace medical measurements. There is still too great a tendency to look upon tests and diagnoses as reliable in an absolute sense.

Beginning with a report by Richard Cabot, a number of comparisons have been made between antemortem clinical diagnoses on hospitalized patients and the pathologist's final verdict. Most of these studies have been done in large teaching hospitals, with good laboratory facilities and readily available consultant services, and the diagnoses should therefore be better than those made in the patient's home or the physician's office, on the average. Yet sizable discrepancies have generally been found. One reason for the imperfect state of the science of diagnosis lies in the variability of symptoms, physical signs, and laboratory findings even in full-blown disease processes. Another may be termed the measurement error; that is, the resultant of all errors in making physical and laboratory tests,

and then interpreting their results. A third source of confusion is a result of the fact that most diseases manifest themselves in a continuous range of severity or extent going all the way from an unrecognizable or subclinical level, on to a maximal severity which may be incompatible with life. This range is sometimes referred to as the spectrum of clinical severity. In an individual the disease may progress with the passage of time all the way from the subclinical level on to death, or it may never go beyond some point part way along on the scale, and may then regress. In general, the nearer we are to the negative or subclinical end of the scale, the greater the difficulty is in making a diagnosis. In epidemiological or clinical studies it is necessary to decide at what point on the scale, and on what evidence, to classify a person as having the disease. The farther in the direction of increased severity or extent we draw this line, the smaller the error to be expected in our list of persons having the disease; in other words, the greater the specificity of our diagnoses. It must be remembered that the penalty for this increased specificity is usually a loss of sensitivity; that is, failure to list other true cases, less severe or less characteristic. In some types of study a high degree of specificity is to be desired, and in others, high sensitivity will be sought, but a middle ground is generally most satisfactory.

The question, how reliable are the diagnoses with which we have to work, must be prefaced with another—how is it possible to validate such diagnoses? This will, of course, depend on the disease. Only in rare instances do we have a single test that will do this for all cases, even all cases that are well out on the scale of severity. In diagnosing diabetes the sugar tolerance test is extremely helpful, but even here the borderline between negative and positive is not clear-cut. The demonstration of virulent tubercle bacilli by appropriate techniques proves the existence of tuberculosis, but significant tuberculous disease is present in many more persons than are sputum-positive; that is, the test is not a sensitive one.

The tissue pathologist can give us reliable diagnoses in many

instances, though he has perhaps been undeservedly immune from criticism up to date. His microscopic diagnoses may and should be tested as to their consistency and reproducibility, like those of other specialists. Assuming for the moment, however, that the pathologist is always right, it is still no solution to our problem to accept only diagnoses that have been pathologically established. The reason is that this involves a high degree of selection of the clinical material, even greater than limitation of a study to hospitalized cases. Suppose that we wish to study the geographic distribution of a series of cases of sarcoidosis; if we require a biopsy before including a case in the series, we shall first bias the types of cases included, because they must all have accessible lesions and second, communities having hospitals which for one or another reason do many biopsies will appear to have a high prevalence of sarcoid. For the vast majority of chronic diseases, then, there is no single pathognomonic symptom, physical sign, or laboratory test. We must depend upon a combination of findings which, taken in relation with the course of disease, is termed the "clinical picture," and which may, in the clinician's opinion, be more or less characteristic. How can we proceed to evaluate the likelihood of error in such diseases, and to reduce the number of errors?

One useful though limited approach, employed in the x-ray studies previously mentioned, is to test the reproducibility of our classification of cases, first by the same individual who made the initial classification, then by a second expert. In diseases where the classification depends wholly on examination of the patient, this is rather awkward to carry out and the examiner is likely to be influenced by his previous decision. He may sometimes be asked to review his written record of earlier findings (for instance, the history and neurological findings in certain neurological diseases) and render a second opinion. Then only the cases on whom there is diagnostic agreement may be selected for study. These techniques were employed by Westlund in a study to be reported at this meeting.

An attitude of skepticism on the part of the investigator is

always justified regardless of the reputation of the medical center in which the patients were diagnosed. An effort must be made to set down the important objective diagnostic criteria and find out what combinations of these are considered essential for a positive diagnosis. It is important, however, that these criteria should not be such as to force the cases into a mold formed by the preconceived opinions of the clinician. As a rather farfetched example of this, if the clinician has a preconceived belief that white females under 50 years of age never have osteoarthritis, and uses that belief in his classification of patients, the distribution of cases by age, sex, and race revealed in his series of osteoarthritics cannot be said to be characteristic of the disease.

Improvements in diagnosis are the direct responsibility of the clinician, to whom they are of the greatest interest and concern. The epidemiologist and statistician, however, need to understand the clinician's problems in order to work most effectively with the data which he supplies them. Frequently they in turn can aid the clinician in the process of analysing and interpreting his material.

To turn from the problems of classification in small, intensively studied series of cases to the analysis of mass data, the only point that will be mentioned here is that the investigator must be familiar with the nature of the raw data and the steps employed in processing it. It is not too much, for example, to ask him to examine at least a sample of death certificates for cases of a disease before making a study of trends of mortality in this disease based upon annual reports of mortality. He may find "jokers" of many kinds, some inherent in the certificates themselves, some introduced in the processing of the certificates. Thus Kurland and Moriyama found that about 18 per cent of death certificates coded and classified as "multiple sclerosis" actually showed "cerebral sclerosis" as the cause of death and that in nearly all these cases the physician meant "cerebral arteriosclerosis," not multiple sclerosis.

If a determined effort has been made to measure errors of

classification, the investigator may become so discouraged by their magnitude that he is tempted to abandon the study. At this point, however, it is important to avoid undue pessimism. No classification of illnesses or deaths can be perfect, better data frequently are unobtainable, yet important information is often available in crude data. Having recognized the imperfections of a classification, even a diagnostic error of 15 or 20 per cent oftentimes need not deter one from going on with a study. Important real characteristics of the material will not be obscured by errors of this size, although they will be less evident than in a "pure" sample of the disease, and while spurious characteristics may be introduced, we must take that chance. Certainly, to demand high specificity in diagnosis where that necessitates a marked bias in the selection of material is unwise.