

EPIDEMIOLOGICAL METHODS IN THE STUDY OF MENTAL DISORDERS

D.D. REID, M.D., D.Sc., M.R.C.P.

*Professor of Epidemiology,
London School of Hygiene and Tropical Medicine*



WORLD HEALTH ORGANIZATION

PALAIS DES NATIONS

GENEVA

1960

PRINTED IN SWITZERLAND

CONTENTS

	Page
Foreword	7
The epidemiological approach	8
The use of vital statistics	17
The measurement of levels of mental morbidity . .	26
Prevalence surveys of mental disorders	36
The measurement of personal predisposition	47
Assessing genetic and environmental factors in the distribution of mental disorder	55
Technical aspects of procedures in field observation .	59
Epidemiological experiments in disease control . . .	71
Postscript	75
References	77

FOREWORD

This paper was prepared as an introduction to ideas and methods of epidemiological inquiry that may prove useful in the study of mental disorder. It is not a text-book of epidemiology; the aim in mind is rather to present, by precept and example, the general principles of this approach in medicine and to point out both its potential and the practical limitations inherent in its application in psychiatry. No comprehensive review of the relevant literature is attempted. The examples cited are chosen simply to illustrate the practical problems most often encountered in the design, conduct, and interpretation of epidemiological studies of mental disorder. By their discussion, I hope to spare the novice in the field some of the disappointments and frustrations which might prevent his going on to do useful work. I have tried to point out to the psychiatrist interested in exploiting any opportunities he may have for such studies how he may tap the accumulated experience of the pioneers in this field and where expert help may be required.

At the same time, this statement of aims and methods may serve as a canon of epidemiological practice which could command the agreement of experienced workers and thus lead to a closer comparability between studies in different parts of the world. The clinicians for whom it is primarily intended will need no reminder of the complications of diagnostic classification in psychiatry. Because of their fundamental importance in the ascertainment of "cases" in epidemiological studies, they deserve a separate account; in this text the problem is only skirted.

In the early stages of drafting this memorandum Professor Jan Böök, Dr Ernst Gruenberg and Dr Eduardo Krapf gave me the necessary stimulus and guidance; and I am most grateful for the informed and helpful suggestions made by the members of the Joint Technical Meeting on Epidemiological Method in Psychiatry held in London in September 1958.

THE EPIDEMIOLOGICAL APPROACH

Epidemiology has been variously defined. Greenwood's description of it as "the study of the mass aspects of disease" had the merit of simplicity and the implication of a wide generality in application. In medicine in general, and in psychiatry in particular, the limitations imposed by a preoccupation with the individual patient unrelated to his social or physical environment have become increasingly obvious in recent years (Milbank Memorial Fund, 1950). Since many mental illnesses are as much "crowd diseases" as is typhoid fever, the epidemiological method evolved for the study of infectious disease applies equally in the investigation of the individual's reaction to his surroundings and of the varied patterns of mental disorder in different or changing social or other conditions.

In psychiatry, as in other branches of medicine, epidemiological inquiry is designed to measure the risk of attack by specific disorders within communities and to uncover clues about their origin and mode of spread. These clues are gleaned from the distribution of disease in relation to time, space, or the distinguishing characteristics of the individuals or social groupings affected. Once the natural history of the disease in a particular population has been established, the epidemiologist is concerned to devise measures of disease prevention or control and to assess their efficacy in practice. Successful prophylaxis achieved by modifying some essential cause is the ultimate test of the epidemiological approach; and a creditable series of such successes has made epidemiology the basic science in preventive medicine.

Comparison and correlation

The first step in this system of disease investigation is a process of description whose essential feature is the comparison of the incidence of a specific abnormality in sub-groups of human populations. This implies the observation of such groups over a period of time, noting the rate of onset of a defined syndrome among people differing in environ-

mental circumstance and personal habits or characteristics. In practice, however, the comparisons must often be restricted to the prevalence of the disease in particular groups, i.e., the number of cases existing in the population at one point in time. Whatever the technical method, the aim is identical: to see whether the risk of becoming ill is greater among groups of individuals with some characteristic, personal or environmental, in common than in others not exposed to the specific environmental circumstance or without these particular attributes. In the same way, changing risks of illness can be related to the experience of similar social or other groups living in different places, under different circumstances, and at different times. As epidemiology can be usefully regarded as a medical branch of human ecology, mental disorder incidence rates can be also correlated with the other indicators of present or past conditions and of changes in population structure or social organization which are used in the study of the relation of man to his surroundings.

The range of investigative method

Epidemiology uses research techniques which range over a spectrum from the almost fortuitous observation of coincident disease and circumstance to the deliberate experiment touching the lives and health of human beings. The emphasis of most of the work in exploratory studies on disease etiology is, however, on observation rather than on experiment. Only when a novel observation leads through a confirmatory planned survey to the framing of a hypothesis which can be put to the test is there scope for the decisive experiment in disease prevention or control. The steps in scientific procedure which form the subject of this text may be briefly outlined.

The initial stimulus in epidemiological research often comes either from a clinician's perception of an apparently unique coincidence of specific disease with a particular social or other characteristic, or from the discovery in an analysis of routinely collected vital statistical data of either an excessive or an unusually low incidence of disease in a particular population group. An example from the problems of psychiatry in war medicine may clarify the ensuing process of investigation.

During the First World War illness presenting with signs and symptoms of cardiovascular and psychic dysfunction among aviators was frequently described as "flying stress"—a disease *sui generis* which was ascribed, at least in part, to the effects on the central nervous system of physical stresses such as the oxygen lack, cold, and noise then inseparable from combat flying (Symonds, 1943). Both in the First and in the Second World War, men were seen with the signs and symptoms which

had been termed "flying stress" but without a history of prolonged exposure to the physical factors, like oxygen lack, which were specific hazards of flight. Clinical histories, in fact, strongly suggested the importance of purely psychological factors—the anxiety and fear produced by terrifying experience in action or in accidents—on men temperamentally predisposed to neurotic illness.

In order to plan preventive measures it was important to know whether (1) the bulk of the cases arose simply because certain individuals broke down under relatively minor combat stresses (i.e., whether host susceptibility was the all-important factor); (2) the risk of neurosis was related to the intensity of such environmental stresses; (3) there was evidence of cumulative "fatigue", due to prolonged exposure to the physical and emotional stresses involved in operational flying. Clearly, if the first of these hypotheses provided the whole explanation, disease control could come only from better aircrew selection; if the last, limitation of the tour of operational service would be of decisive importance.

Time, places, and persons

After the first hint of an association between two factors, the search for clues to causation become more systematic. As Wade Hampton Frost has pointed out, epidemiology is at this stage based on inductive rather than deductive processes of logic. In other words, the evidence accruing from field observations is circumstantial, in that it may be enough to suggest a causal relationship but cannot give final proof of it. Yet the accumulation of such evidence may be vital in uncovering the sources of disease.

The first step in systematic investigation is the splitting up of the population into fairly homogeneous subgroups, and the comparison of disease incidence in groups which are similar in other respects but unlike in one. Such groups might be homogeneous in factors likely to be relevant, such as age, sex and job, but differing in place of work. It is convenient to make these initial comparisons along three axes of classification—by time, by places, and by persons. In the present context, for example, the incidence or frequency rates of referral for psychiatric opinion because of functional disorder in men, all employed as aircrew, were compared in the different Commands of the Royal Air Force at home or overseas (Great Britain, Air Ministry, 1946). There was a significant downward trend from peak rates in Bomber Command based in the United Kingdom, Middle East Air Force, and Fighter Command to the lowest rates in the home-based training formations. Clearly, the geographical distribution of neurosis was related more to the intensity of the air warfare than to separation from home. Again,

within Bomber Command, the incidence of breakdown in the same squadrons over a period of *time* was more closely related to changes in the risks of becoming a battle casualty than to the actual amount of flying done, i.e., the immediate anxieties of action were apparently more important precipitants of neurosis than the cumulative fatigue of continued activity (Reid, 1948). Similarly, the peak of incidence, early rather than late in the operational career of men exposed to the hazards and stress of combat flying, emphasized the importance of predisposition rather than cumulative fatigue. Finally, the personal element was brought into focus by the comparison of the higher incidence of neurotic illness among air gunners, who were less stringently selected, than among pilots or navigators, who had the more onerous and exacting duties. At this stage, then, the evidence suggested that the immediate effects of operational stresses on men temporarily predisposed to neurotic illness accounted for more of the cases than did cumulative "fatigue". Selection, rather than tour limitation, thus appeared to give a better prospect of an effective reduction in their number.

The determination of cause and effect

Extensive or large-scale surveys are designed to provide clues which can be followed up by more detailed intensive investigations. Possible, and even likely, causes of a specific disease are suggested by these preliminary studies in three ways. The simplest criterion of a cause and effect relationship is that the possible agent and the disease should always be found together; with its converse that if the agent is absent, so is the disease. If, for example, neurosis in aviators were caused by severe psychological stress, the disease should occur at times when severe stress is present or at a consistent interval thereafter; and without stress there should be no illness. This simple proposition must be qualified by ensuring that no other factors likely to be relevant to the onset of the disease must vary at the same time. In other words, comparisons between stressful and non-stressful situations are valid only when all the other factors are held constant. In the present example, the higher incidence in bomber pilots than in fighter pilots has a true etiological significance because, if initial selection has not been specific, both groups are men of the same age doing a similar type of job but in very different circumstances. Finally, it is usual to require that the incidence of the specific disease should vary almost in step with the intensity of the supposed etiological factor. Here the incidence of neurosis varied with the stresses implied in the casualty rates which these bomber crews had to face. As a corollary, it is obvious that where presumed cause immediately precedes possible effect, the inference becomes even clearer.

The concept of multiple causation

The simple logical framework thus outlined is more or less appropriate, e.g., in the epidemiological investigation of infectious diseases in highly susceptible communities where the single causative agent of the disease has an overwhelming effect on all exposed to it; the measles outbreaks in isolated island populations like the Faroes are classic examples. But, as the present example illustrates, in most diseases we are dealing with both the predisposing characteristics of the individual and the precipitating effects of the immediate environment. The assessment of their separate influence, or the interaction between them which determines disease, requires a more sophisticated conceptual model. In practice, any notion of a simple all-or-none type of response is almost always quite unrealistic; and clinically evident disorders must be regarded as the net result of an accumulation of genetic and environmental factors.

This notion of the multifactorial causation of disease is peculiarly appropriate in the insidious course of many chronic physical or mental disorders. In pulmonary tuberculosis, for example, the tubercle bacillus may be the *sine qua non* of disease inception, but environmental conditions and personal diet or habits may play a decisive part in its clinical progress and outcome. In mental disorder, there may be some essential factor in the primary inception of the illness; but its evolution may depend on environmental circumstances, such as the influence of interpersonal relationships, which determine its later course. If such disorder be regarded as the ultimate expression of the interaction between these environmental factors and individuals whose vulnerability may vary widely, the limitation of the investigative technique may be accepted with more patience.

From extensive survey to intensive study

When extensive macroscopic reviews of the detailed distribution of a particular mental disorder have thrown up suggestive clues to the relevant etiological factors, the focus of the investigation is altered. The next step consists in testing the hypothesis of cause and effect thus framed by reversing the direction of the inquiry. Instead of comparing incidence rates in population groups similar in all but one vital respect, we start with the sick, and, by comparing with some non-sick control group, determine whether they differ significantly in the attribute or experience thought to be a causative factor in this disease. If, for example, we wished to test the relevance of harrowing experience or family history of mental illness as precipitating or predisposing factors in the neurotic breakdown of pilots, we should compare their past his-

tories with those of a group of their colleagues matched by subsidiary characteristics such as age and operational exposure.

The techniques and limitations of such retrospective inquiries will be discussed later. It is enough to point out that some of the shortcomings can be minimized by prospective surveys which follow the experience of a whole cohort of individuals who have been independently classified in some way before the period of the exposure or observation begins. Pilots in training, for example, can be classified by psychiatrists according to their estimate of predisposition and then followed through their subsequent operational careers in the same type of combat duty. In this way, the hypothesis of the importance of personal loading to ultimate breakdown may be more rigorously tested than by the more usual *post facto* type of inquiry.

The need for field experiment

However satisfying the results of such intensive studies may be, they are open to the objection that correlation does not necessarily mean causation. The hypothetical cause may be found in constant association with the disease effect only because both are separately related to some known or unknown third factor. Poverty and reactive depression may co-exist in the same patient because both are related to chronic alcoholism. Similarly, where two conditions coincide, one may equally be the cause of the other: the alcoholism may cause the depression and vice versa. Although something can be done by comparing sick and non-sick groups closely matched for all characteristics likely *a priori* to influence the onset of the particular disease, there is always the possibility that some factor, not thus taken into account, may be the vital link in the chain of causation. A severe accident, over-indulgence in alcohol, and neurotic symptoms may co-exist in the same pilot's personal history—yet which causes what?

In circumstances where cause consistently and uniquely precedes effect, etiology may present no problem. Unfortunately, such is seldom the case, and the ultimate court of appeal in many circumstances must be the field trial or experiment. Planned observation and experiment are alike except in one vital particular. The same care is taken in both to ensure that the two groups being compared are similar in all respects likely to be of material importance. The essential difference lies in the element of control exercised over the situation by the experimenter, who can either alter environment and watch results or allocate subjects in a strictly random manner to environments known to differ in particular ways. In either event, the subject plays no part in his disposal and the experiment is thus freed of the personal liabilities or bias which bedevil the interpretation of even the best-planned observation of

naturally occurring events. With human subjects and real-life situations, experiment in the strict sense is always difficult, but not necessarily impossible, even in psychiatry. Clinical trials of new remedies are, of course, now an established part of any research programme in therapeutics. In such trials or clinical experiments the changing state of a group of sick patients before and after treatment is compared with the state in a similar or control group not so treated. In the more usual type of experiment conducted in epidemiological research on the other hand, the sickness experience of a group of individuals who are not sick but have been the subjects of some new form of prophylaxis is contrasted with the experience of a control group when both are exposed (as in poliomyelitis vaccine trials) to the same infective hazard. A psychiatric analogue to such experiments might well form the final stage in an epidemiological study. In the military medicine example already cited, proof of the critical importance of acute anxiety, as opposed to cumulative fatigue, in the precipitation of neurotic illness would have involved the random allocation of men of equal temperamental predisposition to types of combat duty similar in the duration of the flying effort demanded but differing in the hazards to be faced. In less dramatic circumstances, the proof of the validity of hypotheses regarding mental health programmes for the individual or community must eventually come from the type of prophylactic trial now current practice outside the psychiatric field; and the technical problems and methods will be discussed later.

Some alternative approaches

Like the observational sciences of meteorology or astronomy, epidemiology is not entirely dependent on experiment for proofs of theories based on the observation of naturally occurring events. As in these sciences, successful prediction can often resolve doubts or give added confidence to belief. If, for example, experience in one type of operational command or duty had suggested that a casualty rate above a certain level produced a rapid increase in the rate of neurotic breakdown, the generality of such a proposition could be tested by comparing predicted with observed neurosis rates in another field of combat where operational conditions might be very different but the hazard involved identical.

Besides the observational approach, there are three other branches of epidemiological research—theoretical, experimental, and historical. The first attempts to build up a theory of epidemic behaviour of disease in communities were based on mathematical models of the consequences which should follow certain specified assumptions about the contact between infective source and susceptibles in defined crowd or commu-

nity conditions. Although these models were designed to fit the behaviour of infectious diseases, there is no reason why some appropriate forms should not apply to the aspects of "crowd behaviour", such as the dissemination of psychological disorders in human populations. Except rather tangentially in the genetic study of the familial distribution of cases, this is one aspect of the subject which has received little attention.

Experimental epidemiology was originally developed to circumvent the technical and ethical problems of experiments on groups of human beings. The course of an infection introduced into a "village" of mice could, it was hoped, be tracked and described, and the relation of the course of the epidemic to changes made in the environment and living conditions unequivocally assessed (Greenwood et al., 1936). Theories developed on *a priori* assumptions could then be tested against the reality of epidemic behaviour in a community. To some extent, the underlying concept is the same as in experiments on the social behaviour of animal groups. Both methods have their potential but, equally, both share the difficulty of the translation of results from mouse to man. Such experiments in human groups may yet link up with more orthodox epidemiology.

The historical approach has a distinguished ancestry. The ebb and flow of infectious disease which followed war, famine, and the movement of human populations have long been an enthralling topic for epidemiologists, and papers such as those by Haeser (1882) and Hecker (1832) on the epidemics of the dancing manias of the Middle Ages suggest that historical research has some value in the study of psychological disturbances of men and society. In so far as the evidence is anecdotal and interpretation often speculative, this aspect lies somewhat outside the main stream of investigative technique with which we are concerned at present.

The link with genetics

Infections are often family affairs; and so are many of the minor and major mental disorders. Whether familial concentrations of disease are produced by specific genetic factors or whether they are determined mainly by environment is often open to question. Are the children in certain families apparently mentally retarded largely because of their genetic heritage, or because the poverty, cultural rather than material, of their family milieu affects their ability to perform the mental tests on which the diagnosis is partly based? For these reasons, the field of epidemiology in its wide sense traditionally includes the study of the family as a group; and, to this extent at least, it inevitably overlaps with genetics. In community studies, too, no assessment of mental disease incidence is complete without a consideration of questions of population

genetics in the determination of the geographical distribution of the disease.

Interpretation and conclusions

Epidemiological data may thus be drawn in many ways from many sources. Both method and source can vary in quality, so that interpretation of results depends on the weight to be given to each item of evidence. In general, confidence in the value of each item will depend on the rigour of the method used and on the consistency of the evidence with that obtained from other stages in the investigative sequence and from other studies, whether epidemiological or not, on the etiology of the particular disease. The analysis of data from the observation of naturally occurring, and therefore almost inevitably complex, events usually demands the use of quite sophisticated statistical techniques, such as multiple regression analysis. These methods seek to measure the association of two variables, such as domestic overcrowding and suicide rates, after coincident variations in other related factors, such as population density, have been taken into account. Simplicity, rather than complexity, more often distinguishes the crucial planned observation or experiment. Here the statistical techniques are not so difficult.

This conspectus of the field of epidemiology, with particular reference to the study of mental disorders, should emphasize the catholic choice open to an investigator wishing to adopt this approach to the etiological problems of diseases of the mind. All the main factors in the etiology and evolution of such illness—genetic, physical, or psychological—can be investigated by epidemiological methods. On the other hand, the surveys of disease incidence or prevalence which form the basis of many such investigations may give information which is immediately useful in the organization and administration of mental health services. In the same way, the methods of assessing the efficacy of measures of disease control can be used to gauge the functional efficiency of these services. The general principles involved may, however, be most readily illustrated by reference to inquiries into the causes and control of mental disorder. It is with the technical aspects of these methods that the following chapters are concerned.

THE USE OF VITAL STATISTICS

Mental disorders are not usually recorded on the death certificate as the primary cause of death. In suicide, on the other hand, death certificates form a fairly reliable guide to the frequency of such events in any defined population. Between these extremes, mortality rates have a restricted but occasionally useful role in epidemiological studies of particular types of mental disease. Their use is a traditional preamble to all inquiries into the natural history of both infectious and non-infectious illness; and their value is high in such inquiries for several reasons. The completion of a death certificate is a legal obligation, and some check is usually made by the medical statisticians of the central government on the standards of certification of causes of death. Death certificate data on suicide or on anencephaly may give a reasonably accurate indication of the frequency of such events. Perhaps most importantly, these mortality data when classified by age, sex, job, and habitation can be related to known populations based on census estimates of the numbers exposed to risk in each of these subclassifications of the national population.

Methods of analysis

The value of death certificate material in the study of the major psychoses or the psychoneuroses may be very severely limited. In patients who have been treated in hospital for schizophrenia, for example, death might well be certified as due to arteriosclerosis. Nevertheless, the analysis of such data has been a traditional and worth-while preliminary in studies of the epidemiology of chronic diseases affecting other systems of the body. The methods of description and analysis and the principles of interpretation involved in mortality studies are now widely applied in morbidity surveys; since such surveys are being increasingly used in the mental diseases, a review of these methods may be helpful.

In the study of suicide, for example, the comparison of risks, which is the essence of the observational approach in epidemiology, consists

simply in calculating and comparing death-rates. These are defined as the number of deaths from a specified cause per 1000 persons exposed to the risk of dying during a period of time in a particular subgroup of the population. In this fundamental ratio of "cases" to "exposure to risk", the estimation of the latter is frequently imprecise. Two methods of calculating the exposure to risk are commonly used, and both involve the notion that the total exposure combines the number of persons exposed and the duration of their exposure. In vital statistical death-rates, the average number of people alive at the middle of the time period are assumed to have been exposed to the risk of death during the whole of that period. When dealing with large groups of patients, e.g., in industries, this mid-year population can be simply estimated by averaging the population routinely estimated at the beginning and end of the year.

Although appropriate enough for the relatively stable numbers of major population groups, this simple technique may be less effective when dealing with a population, such as the patients in a large mental hospital, whose size is changing erratically over the period of observation. In these circumstances, the total number of months and years lived by each individual who was a member of the group for all or part of the period of observation are added. This gives an "exposure to risk" (usually expressed in terms of the total person-months or person-years of exposure) which can then be used as a sound basis for the calculation of the death-rate.

An extension of this concept of totalling the person-years of exposure is applicable when follow-up observations over longer than a year are being made on closed groups such as large groups of families living in certain conditions. In these studies, whether of mortality or morbidity, the focus of interest is on the relative risks of death or onset of mental illness after the beginning of the period of observation. With the steadily wasting population under review these numbers of deaths, recoveries or relapses will be much affected by the falling numbers exposed to risk. Some systematic allowance must be made for this in the analysis of the results of any follow-up procedure. In practice, a very simple modification of the actuarial life-table is used. Table 1 sets out a hypothetical example of this technique. Suppose we start with 1250 individuals under observation. During the course of the first year of follow-up, 60 die and 40 are lost from observation because they have left the district or for other reasons. The 1150 who survive under review at the end of the first year obviously contribute one year each to the total person-years of exposure for that first year. It is then assumed that the remaining 100 who have been lost to observation, either through death or migration, were exposed to risk for an average of half the period, i.e., for six months.

These 100 thus contribute $100 \times \frac{1}{2} \text{ year} = 50 \text{ person-years}$ of exposure to the grand total. The same adjustment is made for the 1150 survivors and 100 losses in the second and again in the same way for succeeding years of the follow-up.

TABLE 1. THE CALCULATION OF DEATH-RATES IN A CHANGING POPULATION

Year of observation	Initial population	No. dying	No. withdrawn	Total exposure in man-years	Death-rate/1000 man-years
1	1250	60	40	1200	50
2	1150	20	80	1100	18
3	1050	12	68	1010	12

Once the total exposure to risk in each year is available, two courses lie open. If the material is large enough to give reasonably stable rates, the risks of death in each year of follow-up can be calculated as before by expressing the annual number of deaths per 1000 person-years of exposure in the appropriate year. The trend in mortality with duration of exposure in two different sets of circumstances can then be directly compared. More often, however, there is neither enough material for statistical stability nor any very appropriate similar experience available for comparison. In this event, it is often worth while comparing the mortality experience of the particular survey with national or other standard death-rates. Buck (1955), for example, compared the number of deaths from cardiovascular disease among hospital patients with functional psychoses observed over a period of years with the number to be expected on death-rates for the whole Canadian population during the same time. For each age-grouping of patients, the total number of person-years of exposure are worked out as already indicated. Then the age-specific death-rates, e.g., from coronary heart disease, are applied to this total to obtain the total number of deaths from the disease which would have occurred among these patients had they suffered the same coronary mortality experience as the rest of their compatriots. The number of deaths observed is conveniently expressed as a ratio (%) of the number expected at standard rates.

The need for standardization of rates

The differing age and sex distribution of the major mental disorders, such as schizophrenia and manic-depressive psychosis, complicates any simple comparison of death- or incidence-rates in populations of widely divergent age and sex structure (see, for example, Winston, 1931).

Methods of producing comparative rates which summarize the experience of whole populations with differing age-sex structures have been evolved for the study of vital statistics of mortality or morbidity. These have been described in full elsewhere (see, for example, A. B. Hill, 1955) and need only be set out briefly here. Two methods are used in practice.

Direct methods of standardization for age and sex

In international comparisons, the age- and sex-specific death-rates, e.g., the rates for women between the ages of 25 and 34 inclusive for each of the countries, are applied in turn to some common standard population. The number of deaths in each age-group are then summed and expressed as a rate per 1000 of the total standard population. In other words, these new "standardized rates" express the overall mortality risk which each country would have experienced had they both had populations of the same age and sex structure. Such directly standardized rates are commonly used in both international comparisons and in the study of time trends within a nation over a long period when the population structure has been changing considerably.

Indirect methods of standardization

In epidemiological studies of smaller groups, however, indirect methods of standardization are more appropriate, since the smaller numbers involved make any age-specific mortality or morbidity rates calculated from them very subject to random variation. The principle and method involved are most simply illustrated in the standardized mortality ratio used in the study of deaths among occupational groups.

In Table 2, the age- and sex-specific rates in some large standard experience, as in all employed males in the country, are applied to the total number of persons in the corresponding age-group in the particular population being surveyed. This gives the number of deaths to be expected in a population of that size, age, and sex constitution had they experienced the same age- and sex-specific death-rates as the standard population. In the age-group 25-34, for example, 10 151 doctors were exposed to risk for the five-year period 1949-1953, i.e., a total exposure of 50 755 man-years. Had they suffered the same suicide rate as all employed men of their age (80 per 1 000 000 per annum), 4 suicides would have occurred instead of the 16 actually observed. The process is repeated for each age- and sex-group and the total number of expected deaths thus obtained. The total number of deaths actually observed is then usually expressed as a ratio per cent. to the total expected number for the age-range 20-64 years.

TABLE 2. COMPUTATION OF A STANDARDIZED SUICIDE RATIO FOR DOCTORS*

Years	Census population 1951	Registered suicides 1949-1953	National death-rates per 1 000 000 per annum	Expected no. of suicides 1949-1953
20-24	1 053	—	60	—
25-34	10 151	16	80	4
35-44	9 747	17	124	6
45-54	8 290	19	218	9
55-64	4 879	9	324	8
		61		27
$\text{Standard mortality ratio} = \frac{\text{Observed deaths}}{\text{Expected deaths}} \times 100 = \frac{6100}{27} = 226 \text{ per cent.}$				

* Data from Occupational Mortality Supplement, Registrar General, England and Wales, 1958.

In this way, the overall experience of the particular group is compared to the standard, after differences in the age and sex structure of the populations have been taken into account. This indirect form of standardization is mainly used in the study of the occupational or social distribution of death in adult populations, but it can equally serve in the analysis of disabling sickness due to diseases of psychological origin or with psychic manifestations.

Useful as these summary indices of death or sickness may be in the study of long-term trends in disease among populations where the age structure is changing considerably over the period, their use should not be allowed to obscure important features of the data. The age- and sex-specific rates should always be calculated and their pattern inspected. In the example of suicides among doctors, for instance, the excess among this occupational group is most marked in the younger age-groups—suggesting that initial selection rather than occupational conditions determines the unduly high suicide-rate among the profession. At the same time, the actual numbers upon which such a conclusion is based must be detailed, so that their significance can be assessed. A useful rule of thumb is to test whether the observed number of deaths is within sampling limits given by twice the square root of the expected number. In the age-group 45-54, for example, the 19 deaths registered exceed the 9 deaths expected by more than the chance limits of $\pm 2\sqrt{9} = 6$ deaths.

Social and occupational factors in mortality

At the time of a national census, the population is divided *inter alia* according to age, sex, present occupation, and housing conditions and

place of residence. Specific death-rates for each major cause of death can thus be computed by relating deaths, e.g., among male physicians between the age of 25 and 34 living in Wales during the five-year period round the census, to the corresponding census population. The implied equality in the description given by an informant at the time of death certification and the description given at the time of census by the man himself may not be justified, but for many purposes the resulting rates are accurate enough. The problem of differing standards of occupational description in the numerators and denominators derived from different sources, e.g., hospital records and official census, is particularly acute in the interpretation of hospital admission data, and some independent check on official information may be essential (Heasman et al., 1958).

In countries where details of deaths by occupation or by the occupation of the husband of a married woman are available, the social correlations of mortality can be approached in two ways. In the first place, the occupations can be grouped so that men of similar social standing or background and similar standards of personal education and behaviour are grouped together into "social classes". The Registrar General for England and Wales, for example, divides the adult male working population into five such social classes, ranging from the professional groups, whose common characteristic is an extended, usually a university, education, to social class V—the unskilled labourers. Hazards to life and health which are specific to a particular occupation are then obvious in a high standardized mortality ratio for that occupational group when compared either with all employed men or with other occupations in the same social class. The social class gradient in mortality, on the other hand, is attributable to differences between these social classes in characteristics common to all members of the class; and it is only very indirectly linked to the occupations concerned. This trend in the standardized mortality ratios for suicide in the range of social classes from I to V is given in Table 3 (Registrar General, England and Wales, 1958).

TABLE 3. THE SOCIAL CLASS GRADIENT IN SUICIDE

Social class	Standardized mortality ratios 1949-1953	
	M	F (married)
I. Professional	140	163
II. Managerial and higher executive	113	100
III. Skilled workers	89	100
IV. Semi-skilled workers	92	84
V. Unskilled workers	117	83

In general there is an excessive death-rate from suicide in the better-off section of the community. More specifically, the rates for particular occupations like hotel-keepers and domestic servants are high, while the rates for clergymen of the Church of England and railway officials are low; and Sainsbury (1955) has suggested that, quite apart from the general association with social placing, some occupations, like domestic service, suffer from social isolation, while in others, clergy or miners, the individual derives support from the religious convictions or social cohesion of his occupational group.

An alternative approach to the social component in disease is by comparing the experience of married women with that of their husbands; for although their occupations may differ, they share the same domestic environments and social habits. Any marked excess over age-specific expectation in suicide, for example, among doctors or farmers, which is not shared by their wives, is thus likely to be specifically associated with, if not necessarily due to, their occupation. On the other hand, the effect of the social background that consorts have in common is seen in the similarity in the trends in the social class gradient in suicide noted in Table 3. Useful as such summary measures are, standardized rates or ratios may obscure important features of the distribution or changing natural history of disease. In the study of suicide, for example, the more detailed examination of the age- and sex-specific rates will show how, in England and Wales in recent years, the rate for women over the age of 65 has been increasing, while it has fallen among males of the same age.

Trends in age- and sex-specific rates may be even more important in the epidemiology of diseases often described as "psychosomatic". Jennings (1940), for example, has shown how the age incidence of acute peptic perforation has changed from a peak among young adult females at the beginning of this century to a maximum among adult males in more recent years; and this change in age and sex distribution has been reflected in the vital statistical rates of mortality from peptic ulcer.

Spatial contrasts in mortality may also prove informative, not only in disorders ending in suicide but also in mental illness associated with the use of drugs of addiction like alcohol or morphia. In France, for example, studies by Ledermann (1953) and others have shown how the regional distribution of the alcoholic psychoses and other diseases associated with alcoholism corresponds with local data on production and consumption of alcoholic beverages. Here, too, time trends in the ratio of male to female mortality from the diseases usually associated with alcoholism fluctuated with the output of alcoholic drinks in France. International comparisons in mortality, showing a clear relationship between consumption of alcohol and national male death-rates from

diseases like cirrhosis of the liver, indicate the menace of this particular form of drug addiction.

The interpretation of mortality data

The potential value of mortality statistics in conditions involving suicides or the physical and mental complications of drug addiction are obvious enough; and they have the added merit of precision in the relation of the event of death to time, place, and the demographic character of the population involved. But their limitations should be emphasized.

Since these data are necessarily restricted to fatal diseases, they are useful only in a relatively small proportion of mental disorders; and these, e.g., general paresis, tend to be associated with gross pathological changes. This is not to depreciate the important mental effects of injury, congenital anomaly, and infection of the central nervous system, but merely to point out that even in the severest acute depressions suicide is by no means inevitable. In fatal illnesses with both organic and psychological elements, death is thus neither an accurate indicator of the frequency of the underlying psychiatric condition nor of its severity.

Differences in diagnostic standards and conventions are perhaps more troublesome when dealing with sickness data, but such disparities can affect the interpretation of mortality rates. Rising trends in the death-rates from diseases which have been thought by some (Ryle & Russell, 1949) to have a "psychosomatic" origin, such as coronary thrombosis, may equally well be due to changes in diagnostic fashion or steady improvements in diagnostic skill. The usual argument that changes in diagnostic skill should affect the sexes equally may not apply in psychiatric diagnosis; and arguments based on time trends in psychiatric conditions may well be treacherous.

By far the most important potential source of error in interpreting death-rates, however, is neglect of the influence of selection. In this context, the process of dying is, from the viewpoint of death certification, an instantaneous event. But the job, marital status, and place of residence of the individual at that last moment of his life may bear little relation to these personal characteristics at the time of onset of a long and progressively disabling mental disease. The division into social class at the time of census is not a rigid caste system, and movement up or down can reflect either the individual's innate ability and drive or his intellectual deterioration. In the same way, failing earning power can determine changes in residence from one area of a town to another. Again, in the initial selection of a job, both the applicant's interests and the employer's judgement of mental capabilities intervene. Later,

intellectual aptitude or temperamental stability as well as physical capacity determine whether a man becomes an occupational or geographical migrant. Death-rates, whether by job or residence, have to be looked at as an endpoint in a continuous process of selection and interpreted with caution. Thus excessive suicide-rates in the poorer parts of a city need not imply that poverty precipitates suicide. It is possible that the potential suicide has drifted into these depressing surroundings, rather than that he has been depressed by them. Conversely, high suicide-rates in early adult life in certain unattractive occupations suggest either that the inherently susceptible have drifted into them or that the less susceptible have risen to better jobs in the social hierarchy. This sort of hypothesis is only one of several possible alternative explanations of the observed association between disease incidence and social class; but it is typical of the sort of question which can lead on to more detailed controlled studies, e.g., of the occupational or residential histories of individual patients.

THE MEASUREMENT OF LEVELS OF MENTAL MORBIDITY

Mental morbidity is protean in character, varying from the brief isolated episode to a continuous unrelenting deterioration. In the chronic mental disorders, the simpler measures of morbidity used for acute diseases such as infectious fevers are thus not completely adequate, for the onset may be insidious and the course prolonged and either progressive or remittent. From the point of view of assessing needs for hospital care, one must know the number of cases severe enough at any one moment in time to require treatment in hospital. Epidemiologically, however, we are more concerned with the relative frequency of inception of mental disorder in various groupings of the population and with the relationship between such inception and external circumstances or events.

The problems of definition in chronic disease have been fully discussed by Dorn (1957). There are three basic units in the measurement of morbidity—the persons who are ill, the periods or spells of illnesses that they experience, and the duration of these illnesses. The term “spells” will have a meaning which depends on the context: thus it may refer either to a stay in hospital or to a period off work. The frequency of illness can be looked at from two points of view: illnesses commencing during a defined period, or illnesses in existence at any time during a defined period. In the first instance, the usual practice is to define “incidence”, either in terms of persons or of spells, as the rate at which illnesses commence during a defined period among the corresponding population exposed to the risk of doing so. The incidence (attack or inception) rate is thus:

$$\frac{\text{No. of new cases beginning during a defined period of time}}{\text{Average number in a defined population exposed to risk during that time}} \times 1000$$

The precise definition of “new illness or case” will depend on the circumstances of its use. In hospital statistics, it will mean a first admis-

sion, in social security systems a claim for benefit. In health survey work, where a pre-existing illness is discovered, the date of discovery rather than the date of onset will bring the patient into the numerator of the incidence rate as a "new case" in a defined period of time. Analogous arguments hold for the measures of point and period prevalence. Because of past inconsistencies in the usage of these terms, reports of surveys or other studies should give a detailed account of the conventions used.

The point prevalence rate is commonly taken as:

$$\frac{\text{No. of cases ill at one point in time}}{\text{Defined population exposed to risk at that time}} \times 1000$$

As Dorn points out, this rate is not realistic when the period of observation cannot be one instant in time. Where this period is relatively short, however, the conventional point prevalence rate is an adequate measure of the frequency of existing disease in the community at one point in time. The actual level will depend, of course, on the clinical grading of the activity and severity of disease required for acceptance as a "case", e.g., on the need for medical consultation or hospital care. When the period of observation is longer than one day—in the Baltimore study reported by Lemkau (1941) it was as long as a year—the "prevalence" rate is strictly the number of clinically active cases counted at the beginning of the period plus those becoming active during the whole period of observation divided by the average number exposed to risk during that time, or of the population at the midpoint of the period. This may be conveniently referred to as a "period prevalence rate" (General Register Office of England and Wales, 1954).

In mental disorders such as the psychoneuroses, where acute symptoms and occupational disability may occur in the same individual at varying intervals, a clear distinction must be made between "persons" and "attacks" or "spells". In the statistics of social insurance agencies, for example, one person may have several "spells" of absence from work for psychoneurotic illness during the course of the year. In these circumstances, the same conventions in principle are used as before but the numerators are modified thus:

Point prevalence — Number of spells of sickness current at a point in time
(spells)

Period prevalence — Number of spells which are current *at some time* during
(spells) the defined period of observation

The duration of illness, however defined, can be measured or expressed either as (1) the days of illness per person among the population exposed to risk during the defined period; (2) the total days of illness per sick person; or (3) total days of illness per spell of illness.

As before, the definition of the beginning and ending of each individual spell of "illness" will depend on the context. The choice of the measure used will depend on the needs of the particular study.

In epidemiological work the relationship between incidence, prevalence, and duration is of particular importance. The prevalence of a specified disease at any one point in time inevitably depends on the incidence of new illnesses of that type and on their duration. Arithmetically, this can be expressed by noting that when the conditions are fairly stable and the time units of measurement are the same—

$$\text{Point prevalence rate} = \text{incidence rate} \times \text{average duration}$$

In other words, the number of illnesses current at one moment in time reflects both the frequency of occurrence of such illnesses and their duration. Here, the reason for the ending of a period of illness, whether by death or recovery, is immaterial.

However morbidity is measured, the aim of the methods used is to establish the rates at which new cases of mental disorder are occurring in the varying conditions of human populations, and how the evolution of the disease is affected by changing circumstances. In other words, we are trying to describe the pattern of disease, not in static, but in dynamic terms—a concept to be remembered in any discussion of the value of sickness data from any source.

In surveys of chronic disabling disease like schizophrenia, it is often profitable to express the end-result of the accumulation of such disease in any given population in terms of the "life-expectancy" of the disease for any individual born into such a group. This "disease-expectancy" or "morbidity risk" means the likelihood that any individual, who survives long enough to be exposed during the period of risk in life when the particular disease usually arises, will develop the disease. This period of risk varies from disease to disease: in schizophrenia, for example, it is usually taken to be between the ages of 15 and 45. Probably the most widely used method of calculating those morbidity risks is that due to Weinberg (1925). It has the merit of simplicity, as the following hypothetical example will show. From a census-type of survey at one point in time, the age and sex distribution of the population surveyed and the number of mentally ill individuals might be ascertained, with results as in Table 4. It is assumed that all affected individuals can be identified by the diagnostic methods used.

TABLE 4. CALCULATION OF "DISEASE-EXPECTANCY RATE"
OR "MORBID RISK"

Male population surveyed		Schizophrenia cases found = 23
Age structure		Total population surveyed = 5 000
0-14	1 500	less pre-risk-period males = 1 500
15-24	900	
25-34	800	
35-44	700	
45-54	500	less ½ risk-period
55+	600	population (15-44)
		½ (900 + 800 + 700) = 1 200
Total	5 000	Adjusted population at risk 2 300
"Disease-expectancy rate" or "morbidity risk" = $\frac{23}{2\,300} \times 100 = 1.0\%$		

To estimate the total exposure to risk during the period of 15-45 years of age, the number of males who have not yet reached the lower age-limit of 15 years (1500) is subtracted from the total number surveyed (5000), of whom 1100 have already passed through the risk period. The males still within the age-limits at the time of census are assumed to be fairly evenly spread throughout the period and to have been exposed on the average for one half of it. Half of their total (1200) is therefore subtracted from the preceding residual to give an adjusted population exposed to risk of 2300. The "disease-expectancy" or "morbidity risk" is then found by expressing the number of cases of schizophrenia alive in the whole population (23) to the adjusted population as a rate per cent. (1.0%). The whole operation can be summarized by the formula:

$$p^* = \frac{a}{b - (b_0 + \frac{1}{2}b_m)}$$

where p^* = morbidity risk

a = number affected

b = total population surveyed

b_0 = number who have not yet reached the period of manifestation

and b_m = number within specific age-limits

Within limits imposed by the small percentages usually found, the usual standard errors of proportional rates can be used to test the significance of the difference between any two such morbidity risks in different populations.

Although Weinberg's abridged method is simple and reasonably effective in use, it neglects the probability that a census at one point in time will fail to discover cases, e.g., of schizophrenia, who died fairly shortly after onset of the illness. Strömgren (1935) has tried to make some adjustment to the total of cases found to allow for this contingency. Reviewing such adjustments, Böök (1953) points out that schizophrenics have even higher death-rates than the population as a whole and suggests a further method of scaling up the estimate of case-frequency to take this excessive mortality into account. In general, such methods increase disease-expectancy rates without affecting the relative ranking of rates, e.g., from different surveys. Where the data are comprehensive enough to allow it, some such adjustment will, on the whole, lead to a more realistic estimate of disease expectancy in the population group concerned

The use of hospital admission data

Most of the work done on the epidemiology of mental disorder using routinely collected information has been on hospital records. The primary interest and value of such material derives from its importance in administrative planning (Felix & Kramer, 1953). Because of various kinds of selection, hospital admission data have serious limitations from an epidemiological point of view. These sources have been systematically considered by Svendsen (Strömgren, 1950), who divides them into three broad groups—"population", "nosocomial" and "threshold-affecting".

"Population" or demographic factors affecting admission rates are those characteristics of the population served by the hospital which, irrespective of the bed availability in the area, change the number of cases available for admission. These characteristics include the size of the population, or its age and sex structure. "Nosocomial" factors concern the hospital itself. Thus a shortage of beds will lower the admission rate as soon as practitioners in the district realize the difficulty involved. A rapid turnover rate, perhaps as the result of better therapy or improved after-care, will have the opposite result. Yet in neither circumstance will the actual level of disease incidence in the population have altered. The "threshold" for admission results from the delicate balance between the severity of the patient's illness, measured in terms of his social adaptation, the current view of both doctors and the public at large on the degree of domestic or personal upset which is tolerable, and the prospect of effective care in the local hospital. An optimistic viewpoint among the laity and general practitioners will in itself increase pressure for admission to hospital. Distance from hospital or the availability of alternative private services, particularly for the less seriously ill patient, will have the contrary effect. In the same way, new and

effective treatments applicable outside hospital, such as ataraxics, will lessen the demand for hospital beds. Often no complete separation of these several factors is possible, but all trends in reported morbidity rates based on hospital admissions must be assessed in the light of their possible effects. By taking account of these factors, Goldhamer & Marshall (1953) have shown that much of the reported rise in the frequency of functional psychosis in the 15-49 age-groups in the United States could be explained by the changing age-structure of the population. The effect of the public attitude to crime and insanity may be the basis of Penrose's (1939) finding of a negative association between the provision of mental hospitals and of prisons in European countries.

The problem of the diagnosis and assessment of the degree of severity in chronic illness is particularly acute in mental disorder. The patterns of reaction which constitute so much of mental variability vary appreciably in character and intensity over a period of time as a result of the fluctuations, partly within the individual himself and partly in his environment, which determine the threshold of incapacity. In practice, one may have to accept an operational definition of the onset of mental disorder as a disturbance of feeling or behaviour which is disabling enough to cause admission to hospital or an inability to work effectively.

Despite these complexities, hospital admission data have been used effectively to indicate those differing patterns of incidence which are the basis of epidemiological inquiry. During enemy occupation in the last war when hospital conditions were fairly stable, the falling of admission rates was shown in more than one country (Vermeulen, 1943) to be likely to be due to a lessening mental morbidity level at a time of greater national or social cohesion.

It is clear from this discussion that first-admission rates are likely to be most useful when the whole community is served by the hospital system whose records are available. In countries like Norway or the United Kingdom, with highly developed state systems of medical care, the indications of morbidity levels for the more severe psychoses given by admission rates are likely to be fairly accurate. Ødegaard (1952) has used such material to suggest that there are differences between the mental morbidity risks among different ethnic groups within Norway, while the Registrar General for England and Wales (Logan, 1956) has produced evidence of much higher admission rates among the lower social classes of that country. In such circumstances of fairly complete coverage, Ødegaard believes that the disparities found between estimates of the risk of mental illness based on special survey and hospital admission data are less than might at first appear. In Norway, at least, many of the mild but slowly progressive cases found at survey thus appear ultimately as first admissions to hospital.

Other sources of morbidity data

The epidemiologist in search of material should certainly not be confined, however, to hospital admission data. In the health insurance plans, social security, or industrial sickness benefit schemes of private industry lies a great wealth of information on the less severe but none the less important mental illnesses and on physical disease where an emotional basis has been suspected. Medical certificates of incapacity may be clinically imprecise, and at times misleading; but in the initial stages of an epidemiological survey they form a potential source of suggestive correlations between disease incidence, personal characteristics, and the working environment. Because the date of onset of illness severe enough to cause absence from work can be accurately timed, the relation of that illness to changing environment can be readily appreciated. Again, one usually has very complete information about the size, age and sex structure, and financial status of the working and other groups exposed to risk, so that age, sex, and other specific incidence rates can be readily computed.

Sickness data from industrial or insurance sources have, of course, much the same limitations as the information on death certificates, although the two sets of information may have complementary assets and deficiencies. They are both influenced by selection, either initial, at the time of taking up a specific job, or subsequent, when the mental or physical stresses of such employment eliminate the less resilient from the more arduous tasks. The frequency and duration of absence from work do give some indication of the incidence and severity of disease in different occupational groups, but the precision of such comparisons is inevitably affected by the differing degrees of disablement illness of equal severity implies for people doing different jobs. The managing director, for example, will be more disabled by a minor brain injury than the labourer. When the sickness record for the period of the individual's working life is available, however, some idea can be gained of the serial responses he makes in terms of sickness absence to the responsibilities and stresses of his career. Similarly, when chronic illness becomes severe enough to demand retirement from work, registers of such cases occurring in different groups may be used to calculate ill-health retirement rates, e.g., from frankly psychological disorders or from presumptive "psychosomatic" illness in relation to work and the working environment. In these ways, morbidity records can give a changing picture of the dynamic interactions involved and illuminate the problem of the natural history of psychiatric disorder in the working population. As already noted, mortality data deal only with an end-point in this history, so that the two sources of information are complementary.

The chief deficiency of the sickness record, compared with hospital notes, lies in its greater diagnostic imprecision. Dealing as they must with the vaguer, less severe manifestations of mental disorder, the certifying physicians are inevitably forced to use symptomatic labels rather than well-defined diagnostic categories. In the study of these numerically important disorders, these descriptive labels need not be disadvantageous. For although the mode of presentation may change, it may be possible to discern broad trends in time of groups of illnesses similar in origin but varying in mode of expression. May Smith (1936) and others showed, for example, how a decline in neurosis was accompanied by a compensatory rise of the same extent in the incidence of gastric upsets in a factory population. As Russell Fraser's (1947) study of factory workers has shown, the indications of the frequency of absence from work due to neurotic illnesses of varying kinds given by the sickness absence record can be confirmed by clinical examination of samples of workers in different industrial environments.

The use of social insurance sickness data in framing hypotheses about the origin of psychosomatic illness is well illustrated by Halliday's (1948) synthesis of the experience of Scottish miners. It also shows how sickness experience can be related to other indices of social behaviour, whether biological or economic. In the interwar years, sickness claim rates for Scottish miners were in general twice as high as for other employed men; and the excess was greatest, not only in accidents and skin disease, but in syndromes labelled as "rheumatism", anaemia, gastritis, and disordered action of the heart and "tachycardia". This excess was also most marked among the youngest men—an age notoriously susceptible to a changing environment. "Epidemics" of nystagmus, or claims for registration as disabled from it, were noted after pit disasters or large-scale dismissals of men. Absenteeism also increased in these circumstances, and with it strikes; and work output fell. Biological indications came from a falling birth-rate among the families of the skilled hewers and getters and an excessive death-rate among their wives, particularly from the hypertensive cardiovascular diseases—an experience curiously similar to that of the wives of skilled textile workers at the same time. Some at least of these findings could be explained by the selective migration of fitter young men from the mining communities, but there is a consistency in the accumulating evidence which underlines the challenging questions about what Halliday has called "a sick society".

Correlation studies in mental disorder epidemiology

In the exploratory stage of epidemiological inquiry, we seek relationships between disease incidence and environmental or other circumstance. Unfortunately, because of the intractable nature of the

problem of definition, either of cases or of intangible external conditions, such correlations in the field of mental disorders are often difficult either to elicit or to interpret.

The effects, if any, of the social and physical environment are usually studied by comparing the incidence of disease either in defined geographical areas or in specified social or other groups, where the conditions prevailing in such areas or groups can be expressed in some quantitative form. These groupings must be small enough to be reasonably homogeneous in character but large enough to provide stable rates. Thus Faris & Dunham (1939) found high admission rates for psychoses in areas of cities where indices of residential mobility were high and where indices of socio-economic status were low. Indices of this general type have long been used to measure the correlation, e.g., of physical diseases like pulmonary tuberculosis with physical conditions such as domestic overcrowding, as measured by the proportion of people in the particular area living more than two to a room. Similar indices can be constructed to measure economic conditions (e.g., mean annual income) or social mobility and social stability (proportion of population born in area, or relative frequency of children coming before the juvenile courts).

Individuals can also be allocated to social groups, perhaps on grounds of occupation, type of residence, and education, and these groupings ranked according to some arbitrary scale of social status (Hollingshead & Redlich, 1958). When such groupings, geographical or social, can be assigned some numerical rating on a scale of assessment, the relationships between social conditions or status and the risk of mental disorder can be more fully explored by sophisticated types of multiple regression analysis. By their use it is possible to express in numerical terms the closeness of relationship between disease incidence and social index after adjusting for the simultaneous variation of other factors likely to be relevant. The technicalities of the statistical methods involved can be found in any textbook of statistics (e.g., Yule & Kendall, 1950) but their practical application usually needs a special competence in this aspect of social statistics.

The assumptions underlying the use of geographical or area comparisons in this approach to the epidemiology of mental disease have been critically discussed by Robinson (1950) and Clausen & Kohn (1954). It cannot be assumed that, because disease incidence rates are high in areas with a large proportion of overcrowded homes, the rates would be found to be higher among individuals living in more crowded conditions than among those from less crowded homes, if such a comparison were possible. In other words, although the rate of admission to hospital for schizophrenics may be highest in areas with much domestic overcrowding, these patients may not come most frequently from over-

crowded homes. As in the Faris & Dunham example, the interpretation of the findings may thus be obscure, but such studies have their greatest value in suggesting hypotheses of cause and effect which can be put to the test of more detailed field inquiries.

PREVALENCE SURVEYS OF MENTAL DISORDERS

In any human population, the frequency of mental disorder of recognizable severity at any moment in time inevitably depends on its rate of inception, recovery, and recurrence, and on the mortality among those afflicted by it. Of these three factors, inception is the most relevant in epidemiological studies of cause; and morbidity records are considered of prime value because they reflect the onset of socially or occupationally important disease. Too often, however, such data are either absent or inadequate, and the only alternative is a survey of the current prevalence of disease within the community.

Prevalence studies are designed to estimate the number of patients with a mental disorder of defined severity in a population at one point in time. In practice, this usually means the number counted, not at one instant, but at any time over the whole period of the census survey. Quite apart from any value such surveys have for medical administration, they can therefore be useful in indicating areas of high and low prevalence within the community. Since, unlike routinely collected hospital admission or other data, information is usually obtained by skilled investigators, much more comprehensive studies can be planned. The diagnostic standards for inclusion can be pre-arranged, ancillary methods of assessment used, and much more collateral evidence on personal characteristics and social background collected. On the other hand, there are both considerable practical difficulties in execution and problems in interpretation which must be realistically faced before such studies are undertaken. The objective of an epidemiological survey is to establish disease inception rates for every grouping of the whole population. This is an ideal seldom attained in practice; but it is convenient to consider how far, and for what reasons, various alternatives fall short of it.

As already noted, prevalence at one point in time depends on both the incidence of new cases of disease and the duration of such illness in the individuals concerned. Under reasonably stable conditions, know-

ledge of the average duration of such illness will allow an estimate to be made of the incidence rate. In chronic diseases associated with an appreciable fatality rate, which can vary widely according to treatment and environmental circumstances, however, the relationship between point prevalence and incidence may not be so simple. Yet the approach through census or prevalence studies depends for its validity on the assumption that current prevalence rates do reflect past experience of mental disorder within the particular population group. When the mental disease concerned is incurable but is not associated with any appreciable increase in mortality over normal expectation, this assumption is clearly reasonable. It does not hold for either the minor transient psychological disturbances seen so much in general practice or the severe defect which carries the risk of early death.

The contrast between prevalence and incidence is particularly obscure in cyclical or relapsing types of mental disorders where a susceptible individual alternates between periods of disability and of comparatively satisfactory adjustment. It is in these circumstances that a prevalence study which includes the taking of personal histories covering a longer period of time may give a fairer indication of the frequency in populations of such individuals than, for example, admission data from hospitals which record repeated admissions without relating these admissions to particular individuals. Although the optimum in prevalence surveys is the coverage of the whole population, the cost and effort involved usually makes this impracticable. Sampling of one kind or another thus becomes inevitable, and the essential value of the results obtained will depend on the representativeness of the sample of the population selected and the completeness and efficiency with which this sample is traced and assessed. Broadly speaking, sampling may attempt to cover the whole population, or the survey may be restricted to certain readily accessible groups. Whatever the method chosen, its demerits and assets must be shrewdly weighed with these several sources of bias clearly in mind. The various alternatives may be best set out by a brief review of the survey methods which have been used in the past.

The development of survey technique

The interests of early Swiss and German workers in this field were primarily genetic; and, under Kraepelin's influence, comparisons were made, for example by Jost (1896) and Koller (1895), between the frequency of insanity in the families of hospital patients and in the relatives of healthy individuals. From the point of view of scientific methodology, these and similar studies were not very rigorous. Not only were all mental and neurological conditions rather indiscriminately grouped together to give the total load of hereditary "tainting" by the neuro-

pathic disposition, but there was no attempt at age standardization or at tabulating results by closeness of relationship. Further, the rather casually selected healthy comparative group was very unlikely to be representative of the population as a whole. These defects of method were critically examined by members of Rudin's school at Munich like Luxenburger (1928) and Schulz (1936). They developed a firmer theoretical basis which served to further new work on the familial distribution of mental disease. In essence, the objective was the comparison of the frequency of mental illness among relatives of varying degrees of closeness to a randomly selected group of mental hospital patients with the frequency among corresponding relatives of a representative group drawn from the presumably healthy population. Practice fell somewhat short of this theoretical conception, and practically all the presumptive random samples of the general "healthy" population were approached because they were readily accessible through some connexion with the hospital or the patients. They could not, therefore, be randomly selected in any modern sense of the phrase. The numbers surveyed were quite small, and the rates based on these numbers inevitably subject to serious errors due to chance variations. The desire for expectancy rates in the general population was unsatisfied and prompted other field studies of which one by Brugger is typical.

In 1929 Brugger (1933) conducted one of a series of such surveys in one of the administrative districts of Thuringia in Germany. This survey coincided in time with a census carried out on behalf of the Ministry of Finance. Case-finding used several sources of information, beginning with all patients in the local mental hospitals who came from the district, but excluding those who had come for treatment there from outside it. Cases were also found by asking doctors, clergymen, teachers, and mayors to register patients personally known to them. The oldest inhabitants of villages were questioned, and psychiatric examinations were carried out on families presumed to be normal. At the same time, careful family histories were taken from all patients admitted from the district to the local mental hospital during a period of two years. Using all these case-finding resources, Brugger was able to estimate the total prevalence of mental disorder in this population to be 1.31%—a result he thought consistent with previous findings, although any very direct comparison was, for several reasons, impossible. Quite apart from differences in the age and sex structure of populations, much depended on the category of mental defect or disturbance included, and there were practical difficulties in ensuring a complete census of all afflicted people. Teachers, for example, were most unwilling to give information about their neighbours, while the members of families already known to possess one affected member were particularly helpful in this respect. Because

of presumed differences in the inherent mental stability of migrant populations, Brugger eliminated all immigrants to the district from his survey.

A critique of earlier methods

This and other early population studies were progressive and important steps in such an approach to mental disorder. But they were subject to certain criticisms, directed against the smallness and unrepresentative nature of the population actually examined and the resultant insecurity of the conclusions derived from the results.

A rigorous approach to sampling theory and practice is a fairly recent development in field studies in medicine. It is hardly surprising, therefore, that these earlier investigators appeared to be less concerned with the sampling variation of prevalence rates based on relatively small survey populations than we might be today. More important than chance variation in calculated prevalence rates, however, was the consistent bias introduced by errors in population sampling procedures; for, although the theoretical needs were well recognized, e.g., by Luxemburger (1928), the practical applications of sampling theory were less effective. The crux of modern sampling technique lies in the condition that for the drawing of an unbiased sample each individual in the population must have an equal chance of appearing in that sample. Any sampling scheme which tends to pick out one type of person rather than others is thus useless. As a corollary, it might be said that the selection of one individual is quite independent of the selection of another. In the early studies, these essential propositions were not fulfilled, partly because the emphasis of interest lay on genetic aspects, and partly because limited resources appeared to restrict inquiries to the most readily available rather than the most truly representative samples of the population at large. In practice, these considerations made the choice of mental patients' consorts and their families, or of the families of patients with organic disease, almost inevitable. Clearly, in neither of these cases nor in other "control" groups, such as those notifying child-births (Boeters, 1936) or railway workers (Wolf, 1928), was the essential feature of equality of likelihood of choice from the whole population observed. In the first instance, for example, those notifying childbirth are likely to be young married people; in the second, railway staffs have already been selected, implicitly or explicitly, for intelligence and stability. Thus on demographic or occupational grounds, neither group is likely to have the same incidence or prevalence of mental defect or disorder as the population as a whole. This risk was appreciated by the early workers, who tried to minimize the effects of occupational selection by examining the relatives of such index persons. But even such relatives

cannot be properly taken as randomly selected representatives of the population as a whole. Patients in hospital for diseases other than mental disorders are equally unlikely to be typical of the general population, since infectious diseases like tuberculosis tend to be concentrated in the poorer parts of the community. In the same way, the relatives of such patients have increased likelihood, because of the social environment they have in common, of differing from the rest of the population.

Even when these groups have been chosen for survey, further selective bias enters in the degree of contact made with the chosen population group. The difficulty of seeing and examining all those in a selected group was already noted in Brugger's study, and those who are not seen are notoriously likely to be atypical members of their working and family group. Their hesitation in coming forward for examination may be due to unusually robust resistance to suggestion, or to a reluctance to expose their personal, intellectual, or temperamental shortcomings. In either event, those who are actually questioned cannot be typical representatives of the group. In surveys directed at the discovery of severe disabling disease, this may not be too important, but in less disabling psychoneuroses this process of self-selection may be crucial.

A rather more subtle source of error in the epidemiological interpretation of rates obtained in such "cross-sectional" population studies comes from the time element in the selective factors which result in current prevalence rates of psychiatric morbidity. All groups within a human population, surveyed at one point in time, are survivors from the biological hazards previously encountered in the history of that group from conception onwards. Current rates of morbidity prevalence need not, therefore, necessarily give an adequate measure of the accumulation of risks to be expected among those born into that social or other group over the whole of their life within it. As already noted, inception of mental disorder may be followed either by recovery, complete or partial, or by death. Both recovery- and death-rates are susceptible to social influences through better opportunities for effective treatment or by mitigation of the risk of premature death associated with certain forms of mental disease. Effective therapy may decrease the number of late complications, as in tertiary syphilis, and lead to an under-estimate of the incidence of primary infections. Current prevalence rates among the married and unmarried, or among legitimate and illegitimate children, thus inevitably reflect both the innate qualities of mind and physique which favour marriage and the social milieu associated either with illegitimacy or differences in marital status. In the same way, the phenomenon of social drift—so relevant in occupational or social class differences in mortality from suicide—is an important determinant of

current prevalence rates of psychiatric illness. This dynamic concept of the longitudinal evolution of disease in individuals and groups is of particular importance in the interpretation of prevalence rates at specific ages. Estimates of life-time risks of psychiatric disablement must take into account the period of life when the risk of specific illness usually exists. In schizophrenia, for example, most of the patients first show signs of severe disease between the ages of 15 and 45.

Comprehensive and selective population surveys

These several aspects of the problem of epidemiological estimation of disease risk have been tackled in ways which provide at least partial solutions. One approach strives to minimize sampling variation and, to some extent, consistent bias by ensuring the examination of complete and large groups of the population. The other attempts the direct measurement of disease inception by "longitudinal" studies of the lives of individuals within their social setting.

Comprehensive area surveys

Since Brugger's early studies, perhaps the most successful prevalence surveys in practice have been those which concentrated on the complete coverage, either of discrete communities within the national population or of the whole of the population in one age-group. Of the first type, the studies of Lin (1953), Strömberg (1938) and Bremer (1951) are typical. Lin and his collaborators conducted an intensive psychiatric census of all the 19 931 people living in three areas of Taiwan—a rural district, a small town, and a section of a large city. Bremer described the pattern of mental disease in an isolated Norwegian community where he was district medical officer during the Second World War. In both studies the contact rate was high enough to include practically the whole community population. More recently, a number of similar surveys have been or are being carried out (Felix & Kramer, 1953). They differ somewhat in design and conduct, although they have in common the aim of establishing incidence rates of mental disorders by the observation of populations for varying periods of time. Thus, the Stirling county project directed by Dr Alexander Leighton concentrates upon the assessment of various methods of case-finding and the way in which the results of such methods can be checked by screening tests of the population as a whole. Mental morbidity as thus uncovered can then be related to measures of social stresses within the community. Hollingshead & Redlich (1958) have reported the relative frequencies of patients receiving psychiatric care at the time of a census conducted in different social groupings of the New Haven area. These and similar studies primarily concerned with etiology deal, of course, with single communities which are unlikely to be typical

of the nation as a whole. However suggestive the associations between mental disease incidence and social factors within these areas may be, the estimates of disease incidence thus made cannot be applied with certainty to the United States.

Selected group surveys

The alternative approach through the examination of a complete cross-section at one age is perhaps best seen in studies of intellectual mental abilities and disabilities among Scottish schoolchildren (Scottish Council for Research in Education, 1949), where ascertainment was by the use of a standardized test of reasoning power applied to all children at the age of 11. Such surveys can, as in this example, be repeated at intervals to estimate national time trends in the distribution of intelligence, or be used to determine the frequency of relative defect in different social, geographical, and other groupings of the population. The apparent comprehensiveness of such a survey should not, however, obscure the less obvious limitations. Schoolchildren, although in general a readily accessible group of the population, are none the less selected. They are, for example, the survivors of their generation of births, and those with mental defects like mongolism have been partly removed by the excessive mortality associated with that and other mental disorders. Again, many defective children will not be in school at all, but be cared for either at home or in institutions. In the same way, many children sent away to private or boarding schools will not appear in the ordinary school population. The problems of ascertainment are particularly relevant here. Even when apparently objective standardized tests of intelligence are used, problems of appropriateness to the native culture or of test sophistication arise in any discussion of spatial or temporal contrasts in test performance. Similarly, when psychological upset is being measured in terms of conflict with legal authority, such contrasts will depend for their justice on the completeness and consistency of reporting of delinquency in differing environments. These objections were clearly in the minds of the experienced investigators in charge of these studies; but they may not be so immediately obvious to those new to the field.

In many countries, military conscription affords an obvious opportunity for this cross-sectional approach to morbidity measurement. In the work of Hyde & Kingsley (1944) for example, the prevalence of psychiatric disability was interestingly related to the racial background and present domicile of Boston conscripts. Here again, however, the completeness of the coverage may be, as the authors recognize, more apparent than real. The grossly disabled, e.g., those in hospital with chronic tuberculosis or schizophrenia, may not be called for examination. Workers in essential industries or conscientious objectors on religious

or other grounds may be exempted from military service. Yet these various groups may contain an unusually high or low proportion of the psychologically abnormal. Where the results of routine medical examinations form the basis of estimated prevalence rates, two aspects of this mode of case-ascertainment are important: the fact that the medical examiners are concerned only with the immediate suitability of the candidate for military duties, and the major discrepancies in clinical standards for acceptance which may exist between medical boards in different parts of the country or in different countries (Dahlberg, 1931).

The other type of population fairly readily available for prevalence studies is formed by occupational groupings where medical care and clinical records are supervised by a centralized industrial medical service. The London Transport Executive has published an extended account of the age/sex-specific sickness absence-rates for all trades of workers, drivers, conductors, clerks, employed by that organization (London Transport Executive, 1956). The use of morbidity records from such a source has already been discussed. The main advantage of such groups, e.g., of clerical workers in local or central government offices, lies in the fact that the prevalence of psychological illness in male clerks working in large cities can be contrasted with the experience of their colleagues doing the same job for the same pay but in, say, the smaller towns of the rural areas of the country. In the same way, the prevalence of neurotic illness among male and female clerks, in married and single women working at the same job in the same office at the same time, can be compared. In such comparisons, the psychological influences of incidental occupational differences (social, financial, and environmental) can thus be taken into account. The indications about comparative incidence given by such morbidity analyses must be followed up by clinical surveys, as in the work by Russell Fraser already cited, to confirm or deny the patterns of psychological disorders thus observed.

The longitudinal type of investigation

The difficulties inevitable in prevalence studies are obvious enough and an alternative approach is clearly essential. Klemperer (1933) was the first to appreciate the inherent limitations of the census type of survey and to attempt a radical solution of the problem by carrying out a longitudinal study. His aim was to start with an initial cohort of 1000 births, selected from the registers of Munich between the years 1881 and 1890, and to follow these individuals through their subsequent lives either until 1931 or until their deaths, if these had occurred before that date. In this way, a truly representative, unbiased sample of all born into the Munich population was to be obtained, and inception rates at attained ages computed. Theoretically, this is an ideal method, but Klemperer

found that its practical execution introduced biases in observation, differing somewhat in character, but perhaps as great in degree as those found in the more usual type of prevalence or census study. The difficulties he encountered are of general methodological interest.

The initial selection from the birth registers was not strictly random in the technical sense: samples were drawn separately from the registers of legitimate and illegitimate births. Of the 508 males and 492 females thus chosen, only 271 were found in 1931. Of the original 1000, 524 had died and 205 could not be traced. Local police records had been changed in the meantime, and the card register, which gave the names of only those people who were actually living in Munich, dated from 1920. The destruction of some civic records for individuals with surnames beginning with B, L, G, K, and P in the disturbances in Munich in 1919 limited the choice of the research workers in their follow-up study; and the bias introduced by alphabetical sampling, where family names have both regional and racial associations, e.g., with blood group or other constitutional factors, is well known (Roberts, 1948). The fallibility of the use of surnames in tracing individuals was particularly marked among illegitimate children and young adult females after marriage. One case illustrates both these aspects of the follow-up problem: a girl born out of wedlock, whose mother later married outside Munich and legitimized her birth, herself married and lived in Munich under her new name. It is hardly surprising that she was traced purely by chance, particularly in view of the confusion introduced by the numbers of people in Munich with the commoner German surnames.

In principle, the objective of this study was to reconstruct the life-history of all these 1000 births; in practice, these histories had to be obtained retrospectively, either from records, or from the accounts of their own or their relatives' and contemporaries' lives given in 1931 by the survivors from the earlier generation. This meant that the results were effectively subject to some of the sources of bias inherent in the census type of prevalence study. Between legitimates and illegimates there was the usual disparity in mortality in infant and child life. Quite apart from the tracing difficulties outlined above, there were more survivors among those born in wedlock and thus more information about their psychiatric history during earlier life. For the major adult psychoses this social differential was not perhaps so important; but it illustrates the selective forces operating in the mixture of social and psychological groups in any comprehensive sample of a whole generation. Among the illegitimate births, for example, a high proportion of the mothers had come into Munich to have their babies, or were foreigners or other migrants within Germany. Migration, often an indication of predisposition to, or early stages of, mental disorders, was a major source of

selective loss of subjects from the original sample cohort. Similarly, losses on follow up were greatest among the lower social groups (although no precise breakdown by paternal occupation was practicable). To some extent, the response rate among these groups once they had been traced was also low. For these reasons, the results obtained in this pioneer investigation inevitably fell short of expectations based on the boldness and soundness of its inception.

Essentially the same method has been very effectively used by Fremming (1951) who realized that the advantages of the cohort or generation approach could be reaped only in a small geographically discrete community where migrants could be identified or traced, both inside it and elsewhere, by idiosyncrasies of dialect or family name. The population of the Danish island of Bornholm afforded such an opportunity; and Fremming began with an original cohort of 5500 people born in Bornholm during the period 1883-1887. By using many local sources of information (municipal and parish registers, local doctors, teachers, relatives, and neighbours), as well as by personal visits, he was able to trace 92% of his original cohort. The expectancy rates which he obtained for the major psychoses such as schizophrenia agreed surprisingly well with the figures obtained previously in the same population by means of other methods and with estimates based on the results of census studies in Germany.

Within quite discrete communities, longitudinal studies are thus both practicable and informative; but two considerations may limit the generality of their results. Isolated communities are, by their very isolation, notoriously subject to the pull of large centres of population. In a period of industrial and social development, the selective migration of substantial proportions of the population may make their experience hardly typical of the country as a whole. Again, as Strömberg (1938) has pointed out, the psychiatric morbidity experience of such communities, living in a unique environment where diseases like goitre or cretinism may be common, could be quite unlike experience elsewhere. Isolation has also the effect of increasing the likelihood of consanguineous marriage. However useful this effect may be in studies of genetic influences in the familial and community distribution of mental disease, it further complicates the application of the findings in either epidemiological generalizations or comparisons.

Limited period longitudinal surveys

Some compromise between the theoretical rigour of the longitudinal study and the practicability of the census survey is clearly required. To some extent this is provided by studies of groups or communities which are observed over a longer time than is usual for the conduct of a census

but for much less than the whole life-span. In this way, the inception and evolution of mental disorder in a closed group can be observed over a period of time. Meanwhile, the new recruits, e.g., to an occupational group, and those who leave for psychiatric reasons or otherwise, can be examined at both entry and exit. It is thus possible to obtain true inception or incidence rates (and, conversely, recovery or case fatality rates) for that group over the period of observation. The Baltimore study of the frequency of mental disorder of all grades of severity (Lemkau et al., 1941) is typical of a geographic enclave whose population was observed for one year, while the number of reported episodes of illness or delinquency were noted for each individual. Prevalence rates computed on this basis are, strictly speaking, period prevalence rates, and they cannot be readily compared with the results of studies expressed as point prevalence rates. The component of disease inception, rather than accumulated attacks, has been emphasized in a longitudinal study of a sample of white families in the same district (Downes & Simon, 1954). Here, families living in the Eastern Health District of Baltimore were visited at monthly intervals over periods ranging from three to five years. These families were divided into two groups according to the presence or absence among them of an index case of psychoneurotic disorder, i.e., the first such illness found in that family during the survey. The inception rates of either fresh cases of psychoneurosis or of physical or "psychosomatic" illness among the families with and without an index case over the period of observation could then be readily compared. This method revealed an undue concentration of physical disease, e.g., hypertensive heart disease, as well as of psychiatric disability, in the families where a psychoneurotic index case had been observed.

THE MEASUREMENT OF PERSONAL PREDISPOSITION

At some point in the study of the disease experience of human populations, the problem of the relative importance of differing degrees of personal susceptibility inevitably arises. The incidence of clinically evident stress in any group will depend in part on the resistance to the particular disease among those comprising it; and this resistance may be affected by many factors, innate or environmental. Some discussion of epidemiological methods of assessing the relevance of such factors to personal predisposition may thus be appropriate.

Traditionally, personal liability to disease or injury has been investigated by comparing the frequency distribution in incidents per person observed in any group subject to the same environmental hazards over the same period of time (Greenwood & Yule, 1920) with the distributions to be expected under certain theoretical conditions. The original work led to the concept of "accident-proneness" and a considerable literature has accrued on the psychological qualities presumed to underlie it. In essence the theoretical distributions are based either on Poisson's law of equal individual liability to infrequent chance events or on other theoretical distributions such as the negative binomial, which presupposes that individuals vary in some consistent way in their personal predisposition to accidents. Comparisons between the distributions observed in practice and those based on theoretical expectations give an indication of the hypothesis which agrees most closely with the facts. The mathematical basis of these methods has been fully discussed by Arbous & Kerrich (1951) and the practical implications in the medico-psychological aspects of industrial medicine by Adelstein (1952).

A survey similar to that of Downes & Simon was conducted by Smiley et al. (1955), who extended this notion of the analysis of longitudinal life-histories in the field of mental illness by using the records of a prepaid medical care plan. They studied the question of personal liability to repeated attacks of illness of the same, related, or other kinds by comparing the frequency of fresh attacks of psychosomatic illness

in March among those with and those without similar disorders in January. In the same way, the accident experience of those same groups could be compared to show whether those liable to one type of illness were also susceptible to another. This method is perfectly general in its applicability to such problems. Alternatively, it can be carried out, when the period of observation is longer and several episodes of illness can occur within it, by calculating correlation coefficients between the frequency of specific types of illness in the same individual but in different periods of his life-span.

In all these studies of personal liability, as displayed in life-histories, the individual has survived long enough within the period of observation to have had repeated attacks of the same sort of injury or illness. In more dramatic and dangerous conditions, on the other hand, the first episode or injury may effectively remove the individual from observation. Studies of "accident-proneness" in aircraft pilots have always suffered from the serious and inevitable disability that the pilot's first accident may well be his last. Alternative methods, of approach are thus required.

In any closed community the presence of personal liability as a factor in disease incidence has to be rather indirectly measured by comparing the inception rates among individuals at different stages of experience within any environment. The analogy with infectious illnesses like measles in a class of children, which produce either permanent immunity or death, is complete; for the rest of the period of observation the affected individual is not exposed to the risk of further attack. If all individuals in a group were equally susceptible, the attack rates among new entries and longer-term residents would be substantially the same. But if prolonged exposure to the environment conferred relative immunity, or if the fresh entrants contained a high proportion of susceptible individuals, the inception rate would fall from a peak among the novices to a minimum among the selected and immunized remainder. The parallel in psychiatric sickness is perhaps best seen in neurosis among military personnel in battle conditions. The high neurotic breakdown rates among the relatively inexperienced aircrews in operational squadrons of the Royal Air Force during the last war (Great Britain, 1946) is a typical example of the effects, at least in part, of the increased liability of operationally inexperienced men not previously subject to the selective elimination and immunizing effects of battle conditions.

When long periods of observation are available, e.g., from the records of occupational or general medical practice, the changing pattern of disease susceptibility can be seen by tabulating the age-specific attack rates from particular forms of illness. In the infectious illnesses, like the common cold group, the decline in the attack rate with increasing age

presumably reflects a lessening susceptibility, and the same process may be manifest in the equivalent data on migraine among older people (Fry, 1957).

The coincidence of different diseases

As already remarked, correlation as well as comparison may give worthwhile clues to etiology. Among the most important of associated events giving a clue to causation is the occurrence of different diseases within the same individual, either at the same or at different periods of his life. If the coincidence of two diseases is more frequent than is likely to arise by chance, there may be a reasonable first suspicion of common ground in their separate etiologies, although other explanations should be sought. Such coincidences can be looked for by comparing the frequency of other specific mental and physical disorders in an unselected group of patients suffering from a particular mental disease with the frequency found in the population as a whole. Similarly, the mortality experience from other diseases among patients with severe chronic mental illness may be informative.

Two illustrations of such an approach come from studies by Downes & Simon (1954) and Buck (1955). In the first, the life experience of neurotic individuals (rather than of their families as a group) was followed up, and the frequency of other chronic illnesses noted. This result could then be compared with the number of such illnesses to be expected on the basis of age- and sex-specific experience in the whole of the population included in the contemporary Baltimore survey. The method of calculating the total "man-years" of exposure in the neurotic group has already been described. Essentially the same procedure was followed in the Canadian study of the causes of death among patients in hospital with chronic psychoses. In this instance, however, the standard or comparative experience was the age- and sex-specific death-rates from particular organic diseases in the Canadian population as a whole current during the period of observation. Both studies served to point out the definite association between cardiovascular and mental disorders of both moderate and severe degree in the same patient. In this context, there is thus some evidence of a common origin for both types of disorder.

Two points are relevant in the interpretation of the results of such studies. Conditions, such as coronary thrombosis and duodenal ulcer, may occur together in the same series of personal histories more frequently than might be expected on a chance basis. Although there may be external evidence that both have a psychological element in their causation, this demonstration of their coincidence gives no proof of it. Both may be the results of the organic effects of cigarette-smoking. It is possible, of course, that cigarette-smoking is itself the expression of some

deep-seated psychological malaise, but a more direct and simple physical relationship between the three findings should not be too readily discarded.

In these and similar studies, the selective nature of the clinical material requires emphasis. As Berkson (1946) has pointed out, the apparently excessive coincidence in hospital patients of two conditions at the same time may be a statistical artifact, since admission to hospital for patients with two obvious conditions is statistically commoner than for patients with only one. This excess is additional to the possibility that a patient with physical as well as mental disorder is more likely, on medical or social grounds, to be sent to hospital. The difficulty is less prominent in field surveys where the whole population, or a random sample of it, is included. The need for proper controls in studies of coincident disease based on hospital patients is thus obvious.

Retrospective inquiries

Perhaps the most usual and often the simplest way of assessing the importance of personal factors in the development of mental disease is the retrospective uncovering of personal characteristics, habits, or items in the past history of patients with mental disorder. This is the traditional method of clinical psychiatry, but its scientific validity in epidemiological studies depends entirely on the representativeness of the cases of a specific disorder selected for study and the adequacy of the "normal control" material with which the results of the sick groups are compared. The characteristics—clinical, social, or personal—of patients in a hospital for the chronically sick are likely to differ markedly from those of patients attending an out-patient department. Studies of the frequency of these characteristics in schizophrenics should thus comprise a representative sample of all such cases; or, if restriction is inevitable, its nature should be defined. Without "control" comparisons, unbiased interpretation is impossible. It would be rash to assign significance to a finding that 40% of patients gave a history of nail-biting in youth, without some knowledge of the frequency of such a habit in the population at large from which these patients are drawn. Yet this is a common error in clinical research. The main effort in such inquiries has thus to be put into the collection of a suitable "control" series and the application to it of the identical investigative procedure.

Comparative, or "control", groups are usually selected to represent individuals in the population who have some of the same basic characteristics, e.g., of sex, age or occupation as the sick *propositi*. Although such matching sounds relatively simple, the practical difficulties may be considerable. If the *propositi* are hospital patients, this fact in itself may mean that they are already selected from those in the population

for whom admission to hospital is either fairly easy or acceptable. On the other hand, the selection of "controls" from other hospital patients suffering from other diseases or injuries may offset the selective factor of hospital admission in general, but other incompatibilities may enter into the crucial comparison. Dunbar (1946), for example, suggested that fracture patients selected as a possible control group in a study of the psychological background of cardiac patients themselves betrayed a pattern of psychosomatic experience which would hardly be classed as "normal". Experience in the field of cancer studies has shown how the personal history given by the same type of patient in orthopaedic wards may depend on the length of stay in hospital at the time of inquiry, since the social factors, such as home conditions and family ties, are related to the speed of discharge from hospital. "Controls" chosen from hospital patients may thus be frequently unsatisfactory.

More difficult to find, but more effective, are "controls" selected in some strictly random fashion from the same basic population as the *propositi*. Thus if the *propositi* are psychoneurotic patients seen in general practice, suitable "controls" might be selected by a random sampling procedure from all other patients on the list of the doctor at the same time in the same practice who are of the same age, sex, and socio-economic status, but not necessarily under the doctor's care at the time. Details of the early life-history, of education, and of traumatic events in their recent experience, can then be elicited from both groups, and the frequencies compared.

It should be recognized, of course, that with matched samples such comparisons tell us nothing about the relevance of sex, age, and social status to the inception of psychoneurotic illness. Indeed, the use of such factors in the matching process implies that their importance has already been demonstrated in preliminary wide-scale incidence or prevalence surveys. But such comparisons can be most useful indicators of gross disparities in life experience or personal habits and characteristics which may have etiological significance. Before such disparities are accepted at their face value, however, certain additional sources of bias must be taken into account.

In dealing with mental illness in particular, the very personal and intimate nature of the biographical data may introduce a divergence between *propositi* and "controls". During war, for example, men who have broken down under operational stress are naturally inclined to exaggerate the adversities they have encountered or to admit to previous neurotic behaviour. Their "controls", selected from men who have served in the same combatant units but without breaking down, have no incentive to enlarge on their experience in combat or to dilate on the frequency of mental illness in their family.

Observer bias is of particular importance in the intimate context of psychiatric history-taking, and the personal bias of the interviewer is almost bound to be reflected in the persistence with which he pursues his favourite topics among the sick propositi. For these and similar reasons, the results of retrospective inquiries must be interpreted with caution. The selection of suitable "controls" is perhaps the most difficult part of such studies, and it is here that expert advice is most needed.

Prospective inquiries

Some, but not all, of these theoretical and practical objections to the use of retrospective inquiries can be overcome by the prospective type of inquiry, where individuals who are all presumably well are classified according to certain personal qualities or previous experience and then followed up over a considerable period of time. Any subsequent illness can then be related back to the results of the initial classification. Thus, in a study of family history of mental illness and neurotic disease in later life, the family history, elicited at the initial interview when the subject was still well and had no particular reason to emphasize his family background, would be less biased than a history obtained after the onset of illness. Similarly, events which are recorded in the lives of individuals when any psychological aftermath seemed unlikely are less likely to be over-emphasized.

Whatever their scientific advantages, however, prospective inquiries have their practical demerits. It is usually essential to have some pre-conception of the factors likely to be relevant, and such intelligent anticipation most frequently comes from some clinical impression or theoretical notion which has been confirmed by a retrospective type of inquiry. Even with some forewarning of the likely critical points in personal characteristics or life experience, the numbers of individuals involved enforces a drastic simplification of the detailed data to be elicited and the investigative technique used to elicit them. In the retrospective inquiry, we start with established cases and work back over their life-history in quite intimate detail. But in a prospective study, we have no way of foretelling who of the presumed healthy population surveyed is going to develop mental disease. For most serious disorders the attack rate is so low that large numbers must be surveyed to obtain enough cases to ensure stable and accurate incidence rates in the contrasted subgroups of the survey population. In a follow-up of a disease with an expectancy during the period of observation of 1% it would require the inclusion of at least 3000 individuals in each of the contrasted groups to be reasonably certain of detecting that one incidence rate was twice the other. This means that only routinely recorded data or the results of relatively simple investigative procedures, such as question-

naires administered by lay observers, can usually be employed. It is always possible, if necessary, to increase the likely yield of illnesses by selecting groups known from previous surveys to have a particularly high expectation of illness and to match them with randomly selected "controls" from the same section of the population.

Some of the sources of selective bias already noted in the discussion on longitudinal or cohort studies inevitably reappear in any prospective inquiry where the period of observation is long. Losses from the original body of classified individuals are inevitable; the question is whether these losses are related to the original method of classification. In surveys designed to validate psychiatric opinion on suitability for special employment, for example, it would clearly be nonsensical to expect to obtain a reasonable assessment if men given the most unfavourable prognosis on the basis of past personal history were not in fact allowed to proceed to the ultimate test of that special employment. Yet such mistakes are made. Less grotesque perhaps is the possibility that in a slowly evolving disease, such as schizophrenia, the apparently well may in fact be already ill enough to give responses about their past history which would usually be given by those whose disease is more obvious. In the circumstances, the results of the first period of follow-up must be compared with the results of the later period to identify discrepancies of this general type.

A study of this kind may be cited. One of the main difficulties in determining the relationship between prematurity and subsequent physical and intellectual development lies in the initial selection of subjects for follow up. Children attending either hospital or child welfare clinic are likely to differ in one way or another from all the children of this generation. The solution adopted by Douglas & Blomfield (1958) was to match every child born weighing under $5\frac{1}{2}$ pounds with another above that birth weight, born at the same time, of the same sex and birth rank, coming from a similar home, and with mothers of the same age. In this way it was hoped to equate the biological variables likely to be relevant. Both sets of children were then followed up during their childhood, and their performances measured in terms of age of walking, or their test scores at intelligence-testing in school compared.

In practice, these two methods of inquiry—retrospective and prospective—are complementary, rather than mutually exclusive. Usually they are employed consecutively, with leads given by retrospective surveys followed through in more intensive prospective studies. In the interpretation of the results of both, one must look for sources of bias of varying character as well as for the consistencies in evidence which alone bring conviction. From the practical point of view, the administrative and technical problems involved in keeping a research team

together over a long period of follow up, the large numbers often required to obtain statistically significant results, and the interpretation of results in groups wasted by death or migration make prospective studies hazardous undertakings. They should not be embarked upon without consultation with others who have profited from long experience in their conduct.

ASSESSING GENETIC AND ENVIRONMENTAL FACTORS IN THE DISTRIBUTION OF MENTAL DISORDER

Observations such as those of Penrose (1958) on the wide divergence in the reported rates of anencephaly in different parts of Europe raise the question of environmental as well as genetic elements in their causation. Some of the investigative techniques already described may help to apportion responsibility between these external and internal factors in the inception of specific neuropsychiatric disorders.

Lilienfeld & Pasamanick (1954) have employed retrospective inquiries to elucidate the relevance of brain damage to the foetus *in utero* to the onset of epilepsy in the child. They obtained hospital records of the history in pregnancy of the mothers of 564 epileptic children born between 1935 and 1952 and a similar number of controls matched in respect of race and maternal age. Comparison of these two sets of records showed a higher incidence of complications of pregnancy and prematurity in the histories of the mothers of the epileptic children than in the others. On the other hand, there was no difference in the family histories of epilepsy. On the basis of these and similar results, the authors suggested that these features of foetal environment may be more important in the etiology of epilepsy than genetic factors in the strict sense.

Such studies have the advantage that, in so far as they use already recorded data, they are not affected by the subjective bias of either mother or investigator knowing of the child's defect. Too frequently, however, such records are disappointingly incomplete. "Controls" drawn from a hospital population are also subject to a selection for admission on social as well as medical grounds which varies unpredictably.

Some of the obscurities inherent in hospital samples, whether of defective or of normal children, can be overcome when all the births in a defined area can be observed and their historical characteristics routinely recorded in some detail. From this knowledge of the numbers "exposed to risk", e.g., according to maternal age or parity, the

incidence rates for specified disorders in children born to mothers of varying age and parity can be readily computed. The information thus extracted is, of course, restricted to aspects of maternal history which have been routinely recorded and may be less full than data from hospital, clinical, and post-mortem examinations. Within these limits, however, the series of papers by MacKeown and his colleagues in Birmingham (e.g., MacMahon, 1952) have demonstrated the value of such a reporting mechanism.

Retrospective study is a cheap, quick, and effective method of deciding, for example, whether maternal health, diet, or environment is likely to affect child defect and of indicating the most fruitful lines for prospective inquiry. Such a prospective inquiry was conducted by Mc Intosh et al. (1954), who followed up the children born to 5964 mothers admitted to the ante-partum clinic of a maternity hospital with a duration of pregnancy of 4 months or less on their first visit. Careful histories of the events of pregnancy were taken. The subsequent outcome of the pregnancy—whether abortion, stillbirth, or live birth—was then recorded. Causes of neonatal death and the presence of congenital malformation in the children were then independently assessed. Malformations, many of them major defects of the central nervous system, could thus be related to histories of pregnancy which had been obtained without the personal bias inherent in the interrogation of the mother of a child already known to have some defect.

Family and community isolate studies

Familial concentrations of disease inevitably raise suspicions of a genetic basis, yet epidemiological methods have uncovered other reasons for familial or domestic aggregates of neuropsychiatric disorder. The close time-relation of the crops of cases of psychotic episodes in people, unrelated genetically but sharing the same home, points to an “infectious” spread within a household. Time relationships formed part of the evidence of a dietetic cause of neurological and psychotic symptoms of pellagra. Goldberger (1916) showed how the intervals between cases had none of the regularity of the incubation period of an infectious disease. The seasonal peak coincided with the lean times before the harvest; and women of child-bearing age were most acutely affected. Familial concentrations certainly existed, but among the underfed who might be living in quite sanitary conditions. Except in the rather indirect sense that the families with poorer intellectual endowment tended to fare badly economically, this, then, was a disorder which could lead to psychosis but which was dietetically rather than genetically determined.

While genetic and environmental causes of variation operate in all diseases, it is often difficult, especially in mental disorders, to estimate

their separate contribution. Nevertheless, such analyses are important in epidemiology. When the genetic component is strong and consists of individual single gene differences, one may speak of genetic diseases. The implication is that such mental diseases will occur, following proper environmental stimulation, only in a certain class of genetically distinct individuals; while in others the same stimulation has no such effect. In other words, the predisposition to the disease is discontinuously distributed. The epidemiological features of this predisposition will be determined by the general rules of heredity and population genetics. There are five main research fields in the study of medical genetics:

- (1) identification of genetic disease entities;
- (2) studies of the origin *de novo* of pathological genes, i.e., of mutations;
- (3) identification of the mode of inheritance of pathological genes;
- (4) studies of the effect on the individual of pathological genes, i.e., pathogenesis and biochemical genetics (all genetic diseases are essentially metabolic disorders);
- (5) studies of the behaviour of pathological genes in groups of individuals, i.e., population genetics.

An understanding of these different lines of work is essential for the appreciation of the epidemiology of genetically determined disease.

Especially in the field of mental disease, the study of the effect of pathological genes on the individual may not lead to therapeutically important results in the very near future. But the geneticist is keenly interested in the identification of environmental factors which may inhibit or alleviate the adverse effect of gene mutations. In the broader field of population genetics, the incidence of a genetic disease in a population is determined by the mutation rate, selective migration and fertility, genetic drift (unselective random fluctuations due to the small number of children) and, sometimes, inbreeding.

The traditional way of analysing the genetical and environmental components in mental diseases has been based on twin studies, such as those reported by Kallmann (1946, 1950, 1953) and Slater (1953). In essence, this entails the comparison of the life-expectation of the disease in the twin of an index case of a disorder such as schizophrenia among monozygotic twin pairs, who have the same genes and home environment, and dizygotic twin pairs, who share the same home but are genetically no more alike than ordinary siblings. Still more important is the study of monozygotic twins who have been reared apart. Such data are, however, much more difficult to obtain.

For schizophrenic and manic-depressive psychoses (both of which presumably are heterogeneous groups of diseases) such twin studies have shown a higher expectation of the same type of disorder among monozygotic twins. Although as Slater points out in the introduction to his own study, the detailed explanation of these findings is open to criticism, it seems clear that genetic factors must be taken into account in any analysis of the familial, social, or geographical distribution of psychiatric disease.

The repercussions of genetic concepts on the epidemiological study of disease distribution are implicit in commentaries on population genetics, e.g., by Kilpatrick et al. (1955) and Böök (1953 and 1955). Self-contained communities, isolated perhaps by geography or religion, have obvious practical advantages from the point of view of epidemiological studies. Specific genetical and/or environmental mechanisms may lead to distinct epidemiological patterns; and the components of these patterns are likely to be less difficult to estimate here than in larger and less isolated populations. In small isolated populations, genetic drift may cause large fluctuations in the frequencies of genetic diseases, even if selection against them is strong. There are examples where a local high prevalence rate for schizophrenia is associated with no increase in intra-familial incidence (Böök, 1953). In such circumstances, an environmental or socio-cultural explanation of local variation in disease incidence would have to be made with caution, if at all.

TECHNICAL ASPECTS OF PROCEDURES IN FIELD OBSERVATION

However sound the principles upon which it is based, success in field survey work depends on a meticulous attention to practical detail. Each step in the investigation must be subjected to rigorous testing before use. Some examples of the problems encountered may thus be useful.

Problems in population sampling

The practical advantages in epidemiological comparisons of special subgroups of the population, such as men in specific occupations, have already been pointed out. Inevitably, however, such groups are selected either by personal motives and ambitions or by tests of physical and perhaps mental adequacy for the particular job at the time of recruitment. They cannot therefore be necessarily taken as representative samples of the population at large. Yet to examine the whole population may be beyond one's resources. In these circumstances, sampling is the only practical alternative. Cochrane (1951) has given a useful outline of these procedures in survey work.

Unfortunately, it is often easier to draw a biased sample than a truly representative one. Given a nominal list of the inhabitants of an area, it may be tempting to note that in England, for example, families with names beginning with J, K, and L form exactly 10% of the population, and to use these families as a representative sample of the whole populace. Roberts (1948) has shown how family names, for example in south-western England, are linked with the blood types associated with a Celtic racial background. Similar linkages with diseases like peptic ulcer suggest that, in epidemiological studies in general, serious bias may result from a sampling procedure based on family names. The essential feature of the drawing of a sample representative of any total population is that each individual in that population should have an equal chance of appearing in the sample. Any procedure which favours the inclusion of some or excludes others inevitably leads to a bias whose direction and nature may be unpredictable or unknown but

whose size may be important. In the present context, sampling from names beginning with J-L would exclude the Macs, Owens, and O'Haras whose Celtic race might be related to the risk of particular forms of mental disorder.

To ensure the desired equality in likelihood of inclusion, the sampling procedure must be, if not strictly random, then at least something close to it; at all events it should not be merely haphazard. We all have unconscious preferences for certain numbers and it is thus impracticable to rely on being able to "think up" a random sequence of three-figure numbers. Strictly random sampling methods depend in practice on the use of tables of numbers produced by some entirely random process, such as a roulette wheel, and tested to ensure that no consistent bias is present (Kendall & Babington Smith, 1939; Tippett, 1927; Fisher & Yates, 1938). To draw a random 1% sample from a list of 900 families, the families are numbered from 1 to 900 and the first 9 numbers less than 901 found on going down any three columns of the table are taken to indicate an appropriate sample.

Although, whenever possible, such strictly random sampling methods should be used, the procedure may become laborious when fairly large numbers are involved. In these circumstances, a quasi-random or systematic sampling method may be employed. If, for example, every tenth name is selected from an alphabetical list of persons or families, there is unlikely to be any very serious bias. The whole list must, of course, be covered. Merely to start at A and take every third name until 9 were obtained would repeat the error of nominal sampling by over-representing the first part of the alphabetical list and under-representing the rest. When dealing with military, civil service or industrial groups, where official numbers may have been allocated to each individual, 1 in 10 or 1 in 100 samples may be drawn by taking each man whose number ends, in say, "9" or "59". Where punched cards are available, such sampling can be very conveniently done on the mechanical sorter. It is important, however, that the last digit is used and not the first, for the first may indicate some broad grouping of the people concerned. Whatever sampling method is adopted, its nature should be clearly specified in the report.

Alternative sampling schemes

Sampling is designed to combine economy of effort with the efficient garnering of information about the population at large, and its practical effectiveness can be increased in two ways. Where a research team can cover only a limited area, it may be convenient to sample in two stages. First, the whole district is divided into small areas, homogeneous, for example, in respect of housing, character, and population density, and

a random sample of these areas is drawn. Then from within these areas, a second sample of individuals is randomly selected. In this way, traveling throughout the area is minimized, yet the representativeness of the ultimate sample is maintained.

Where the population can be divided into major divisions or strata, homogeneous in individual character but differing quite widely, perhaps in the presumed incidence of mental disease, stratified sampling may be appropriate. In a civil service, for example, there are relatively few females in the higher administrative grades; yet in a study of the incidence of disorders like psychoneurosis, as an index of the effects of the pressures of government, their experience may be vital. A 1 in 1000 random sample from the whole civil service might well include very few such female administrators. If, however, the service population is first divided into strata (perhaps by sex and grade), random sampling using a high sampling fraction such as 1 in 10 of the female administrative stratum will obtain a reasonably sized group. Their sickness experience can then be compared with a similar sample obtained from a smaller proportion, such as 1 in 1000 of the much more numerous female clerical grade.

This method presupposes that a preliminary census has established the numbers in each stratum. When the incidence rate for the sample from each stratum has been computed, the overall incidence for the whole population can be estimated by averaging the stratum values weighted or multiplied by the stratum populations (Cochrane, 1951).

The accuracy of a sampling procedure depends in turn on the truly representative nature of the sample drawn, and on the completeness of the response from all those selected. A frequent difficulty is the high rate of failure to respond to an appeal to come forward for examination, and the recalcitrant minority are very likely to contain many whose psychiatric history makes them hesitant to report for interview. At a less subtle level, the woman selected in a random sample but not found at home may well be unusual in that her absence may mean that she is concerned more with a job outside the home than with family ties within it. If so, her experience of neurotic illness may differ sharply from that of the house-bound mother of five. Again, a sample which was representative when it was first drawn may be no longer representative if the survey lasts long enough for there to be appreciable shifts in the character of the population sampled. These inherent sources of bias have an importance whose size and direction can seldom, if ever, be accurately estimated. In general, therefore, it is essential to minimize the "drop-out" rate (if possible to less than 5%) and to eschew the insidious attractions of substitution. The woman next door who is conveniently at home is a quite inadequate substitution for the professional woman out at work. In the planning and execution of such surveys, a disproportionate amount of

the effort must go into the seeking out of the recalcitrant minority, and incentives, such as free transport or the hope of a prize in a lottery run among those attending for examination, should be quite seriously considered.

In some branches of social survey work, these difficulties are, to some extent, ignored, and field staff interviewers are allowed more latitude in the final selection of individuals within the particular stratum of the population concerned, e.g., adults in an urban district. In quota sampling of this type, some control is exerted over their choice in that they are instructed to conduct interviews with enough individuals in a clearly specified subgroup of this stratum, e.g., males of the professional class over the age of 45, until their allotted quota has been completed. Within these limits, substitution is allowed. Certainly, the bias thus introduced may be minimized, but its nature cannot be precisely foreseen. The interpretation of results thus obtained is therefore confused by this unknown element, and the economy in field work achieved by "quota" sampling may be outweighed by the unreliability of the results it gives. (For a fuller discussion see Moser, 1958).

Even when a 100% response cannot be achieved, the degree of selection involved in non-response in a randomly selected sample can be roughly assessed from the distribution in the residual 5 or 10% of characteristics not included in the preliminary stratification. Their age and occupational distribution, for example, may indicate the similarity or disparity between them and the more co-operative majority. This majority is likely to be more representative of the population at large than a quota sampled by a less vigorous method, and their sickness experience more typical of the general incidence of disease.

Although consistent bias is the chief source of error, some inaccuracy may stem from the smallness of the sample studied. Where selection has been strictly random, statistical theory allows close estimates to be made of the likely limits of error involved in generalizing about the population as a whole from the particular sample. Quota sampling lacks that important virtue. Random sampling and its modifications thus appear to be the method of choice in surveys of psychological disorder where special circumstances may determine the responses of individuals to the processes of investigation. The general principles of sampling are simple enough, but for the construction of more sophisticated procedures for large-scale sample surveys and in the analysis of their results, statistical advice is usually required.

Preparatory work in population surveys

Field research differs from laboratory work in two main respects: the inability to manipulate environment in a deliberate experiment,

with the consequent reliance on unbiased observation; and the overwhelming importance of the human element in the relationship between field worker and population which such observation entails. The corresponding necessities are thus for a rigorous testing of the methods of observation, and scrupulous care in the enlisting of that population's response and co-operation. Both these aims can be furthered by the careful conduct of a preparatory reconnaissance in the area concerned. In such a reconnaissance, the help of two types of members of the community is vital: those who know about the population's social structure, history, and parochial prejudices, and those whose personal influence is such that they can enlist the helpful co-operation of the rest of the population. Ministers of religion and police officials may be typical of the first group; and labour union leaders exemplify the second in surveys in industrial groups.

In many countries, regular censuses provide quite comprehensive information about the population of localities; and such data may be essential in the early stages of research planning. Broad differences, e.g., in age or sex structure, or in social and domestic conditions, may be relevant to the mental disease experience of contrasting areas. The same detailed account may be useful in constructing a frame of reference for the stratification of the population before the drawing of samples, but important advantages can accrue from the conduct of a special area census before the survey begins. Cochrane (1951) took the opportunity afforded by a private census to explain the purpose of forthcoming surveys and to enlist co-operation; he believes this more personal approach to be more useful than conventional propaganda by press and radio.

Pilot studies, i.e., the conduct of the proposed investigation on a small scale, are often vital to success. They have two main objects: to test the appropriateness of the examination methods contemplated for use, and to obtain estimates of the likely levels of morbidity in the communities being surveyed. When the likely differences and the inherent variability of the method of observation are known, the scale of the final survey required to establish statistical significance for the result can be computed.

The choice of case-finding methods

In the field of mental disorder, case-finding mainly depends on mental examinations carried out by interview or written questionnaire, or by psychological test. Even more care is required in the selection, standardization, and application of such methods than in more usual techniques such as mass miniature radiography. Like any other diagnostic procedure, they have certain qualities which are of particular significance in the comparisons which constitute epidemiological inquiry. Of these, ethical

propriety comes first, for the rules of medical practice in general, and the precept *primum non nocere* in particular, apply with the same force in preventive as in curative medicine. The probing of intimate feelings and relationships by psychiatric interview may bring its own threat to personal and family adjustment, and the framing of a questionnaire must take these susceptibilities into account. Such methods of investigation can be legitimately and effectively used only with the acquiescence of both individuals and community. It should be emphasized, too, that many relevant facts cannot be elicited in the course of a single interview.

Given that a proposed investigative technique is socially acceptable, its application must be within the practical capacity of the research team putting it into effect. This can best be achieved by a simplicity which does not sacrifice the essential qualities of valid discrimination and reliability. A case-finding procedure, whether based on clinical questionnaires or physical measurements, is designed to identify all the individuals suffering from a specific disability of defined severity in the population surveyed. It should pick out all the cases of the prescribed type and degree of severity and include none who fall outside that specification. In other words, these case-finding methods should have the quality of the perfect witness in that they tell the truth, the whole truth, and nothing but the truth. Further, they must go on doing this consistently or reliably over the whole period of the investigation. As in the courts of law, this is an ideal which is seldom attained in practice, but it must be striven after. If this ideal be not attained, at least the observer should know by how much his methods fall short of it.

Some alternative case-finding methods

Hospital admission records have formed the basis of many studies of the incidence of mental disorder, and their value has already been assessed. Information drawn from the examination of military conscripts may be equally useful in certain age-groups, even when used by itself. More can be done, however, by testing the completeness of the indications of mental morbidity levels thus obtained, and by using more than one source of data to gain a better estimate of disease frequency. Thus Dahlberg & Stenberg (1931) tried to estimate the relationship between cases treated in hospital and those remaining outside it by comparing hospital admissions (at all ages) with the frequency of disease amongst conscripts. Srole & Langner (1957), on the other hand, used the results of interviews of a random sample of a New York district to contrast the proportions of emotionally disturbed subjects found in a prevalence study with the incidence of admissions to mental hospitals and clinics. By these methods, they were able to show the difference between social classes in the proportion of the mentally ill who receive in-patient or

out-patient treatment. Carstairs & Brown (1958) combined information from hospitals, police, and other social agencies to obtain an estimate of mental disorder frequency in mining districts of Wales without conducting an inquiry involving personal examinations.

The assessment of validity

A valid case-finding procedure, in this context, is one that is an efficient indicator of the mental condition it purports to identify. In other words, the procedure is a valid screening mechanism for the primary stage of field surveys, where it picks out as affected all those individuals whom some independent mode of assessment registers as mentally disabled. A questionnaire designed to identify individuals suffering from anxiety neurosis, for example, can be validated, at least in this restricted sense, by comparing the results obtained by its use with assessments made independently after a detailed clinical interview by a skilled psychiatrist. Alternatively, if psychiatric infallibility about current and subsequent adjustment to life cannot be assumed, both questionnaire and psychiatrist interviewer can be validated, again in this defined sense, either against records of sickness causing absence from work, performance at school, or additional clinical data from hospital or family physician.

Whichever method of validation is used, it is convenient to compare the results of the initial screening and those of the later method of assessment by setting out the data in the form of a table.

		Result of classification by independent assessment		Total
		Affected	Not affected	
Result of classification by initial screen	Individuals affected	a	c	a + c
	Individuals not affected	$\frac{b}{a + b}$	$\frac{d}{c + d}$	$\frac{b + d}{a + b + c + d}$

Such a table gives a simple summary of the degree and nature of the agreement or disagreement between the classification of individuals by the initial and by the later methods of assessment. Of the $a + c$ individuals, for example, classified as affected by the initial screen, a were also classified as affected by the independent method of assessment. Moreover, d out of the $b + d$ individuals classified as unaffected by the initial screen were similarly classified by the independent method. Agreement between the two methods was thus obtained in $a + d$ out of the total of $a + b + c + d$ individuals; and validity, in a rather general sense, is indicated by the proportion of agreements achieved

out of the total number of examinations. The precise assessment of the technical significance of this degree of agreement is not easy. Two observers may agree purely by chance on the presence in individual subjects of mental conditions or traits, particularly when the trait in question is frequently found in the population surveyed and both observers are aware of its frequency. A purely fortuitous agreement between observers is much more likely when 50% of the population are believed to have the trait than if it occurs in only 10%. Again, the closeness of the agreement in such a two-way classification is likely to be much affected, particularly in less precisely defined mental abnormalities, by the assessors' difficulty in assigning marginal cases to either "affected" or "unaffected" categories.

In assessing the practical utility of a screening or case-finding procedure, two other characteristics are important: *sensitivity* and *specificity*. If the preliminary screen is sensitive, it will pick out a high proportion of those whom independent follow-up shows to be suffering from the disorder in question, i.e., the ratio $\frac{a}{a+b}$ will approach 100%. If the screen is specific, most of those whom it has indicated as unaffected will be found so by the independent assessment, i.e., the ratio $\frac{d}{c+d}$ will be high. The success of the screening device may vary from trait to trait; the review of every item in a questionnaire designed for field use that is related to the three qualities of validity, sensitivity and specificity should therefore be an essential part of the preparatory or pilot stage of a field inquiry.

It may be impossible to interview everyone in the population covered by a survey. The simplest approach then is either a written questionnaire or one usable by a less skilled observer, as a preliminary screen to sift out individuals with obvious disabilities for more detailed examination by specialist members of the research team. At the same time, a continuous check on the adequacy of this screening device should be maintained, by selecting at random for specialist review a sample of those who have undergone the preliminary interrogation and are not considered sufficiently seriously affected to be referred for specialist opinion. The systematic application of such a check should ensure that the preliminary screening does not overlook many cases which a more careful examination would have detected.

The assessment of reliability

Reliability, in the sense of the consistency with which the case-finding method gives the same indication of disease frequency in the same circumstances, requires particular attention. - The sources of variability and bias in the use of any particular method must, therefore, be recogni-

zed, and their size and direction assessed. The stages at which human variability and personal idiosyncrasies of method or judgement may affect the ultimate estimate of disease frequency in groups of subjects can be summarized thus:

Subject variation. From time to time within the period of observation of the same person; consistent differences between subjects.

Observer variation. Variability in opinion in same observer; consistent bias of same observer; differences between different types of observer.

Interaction between subject and observer.

Subject variation

Subject variability is perhaps of most importance in the study of the natural history of physical disease with a possible emotional element, such as essential hypertension. In these studies, measurements of some physical function, such as blood-pressure, may form an essential part of the field observations. The general principles involved could apply equally to assessments of emotional stability. Blood-pressure readings may vary irregularly within the day or over longer periods of observation. Such variations are important in epidemiological work for two reasons. If the fluctuation in blood-pressure is large relative to the average levels observed, the random element in the estimate of the mean of a group of individuals will be correspondingly increased. Greater precision in group comparisons of mean blood-pressure will therefore require the inclusion in the survey of larger numbers of individuals. Again, if there is any consistent variation in time, such as an afternoon rise or a summer fall in blood-pressure, comparisons between one group examined in the morning or the winter will inevitably differ from another surveyed in the afternoon or summer. A vital part of the pilot study is thus the examination of random samples from the same population by the same observer using the same methods but at different times over a period similar to that envisaged for the conduct of the main survey. If sizeable consistent trends emerge, the inherent source of bias thus revealed must be taken into account by allocating observers either at random, or by alternation, or in some other systematic fashion, so that observations are made, for example in the morning or afternoon, with equal frequency in the two or more groups under survey. In this way, comparisons can be made between, for example, the distributions of blood-pressure in these groups, in the secure knowledge that time-determined bias has been equal in both.

Observer variation

Consistent differences between individuals are less important in such group comparisons, except in so far as, like variation within the individual,

they add to the scatter of the observations round the mean and reduce the precision of the estimate based on the sample surveyed. Observer variability, on the other hand, may be crucial in epidemiological studies. As already indicated, it may be of three kinds: consistent bias which is a reflection of the type and intensity of training which the observer may have had; bias which is personal to himself; and the inconsistencies in his own judgement over a period of time which may be either erratic or regular.

Experience in other branches of epidemiology has shown how physicians can disagree, e.g., when presented with the same X-ray evidence, even altering their own interpretations when the same film is seen at long intervals of time (Birkelo et al., 1947). Studies in psychiatric diagnosis (Elkind & Doering, 1925-1934) suggest that a similar divergence of opinions is likely to be important in comparative field studies of mental disorder. Such inconsistencies are inevitable, but as long as their size and direction are known, particularly when they are similar within the subgroups being compared, they are no bar to the use of survey methods.

Two methods of testing observer variation exist in principle. The first is based on the probability that, with complete observer consistency, estimates of disease prevalence in randomly selected samples from the same general population seen independently by different observers should differ by no more than the usual degree of chance variation associated with samples of a given size. The second alternative is the presentation of the same clinical material to different observers or to the same observer at different times. With X-rays or electrocardiographic records, the second alternative is clearly the method of choice, since it gives some indication both of the difference between observers and their overall estimates of prevalence in the groups surveyed and of the degree of their agreement about individual cases. With such information the nature of the disagreement between observers can be studied much more intensively. In psychiatry, however, the intimate nature of the interpersonal relationship at interview makes it difficult to present the same patient successively to two psychiatrists with complete confidence that the patient's responses to questioning have been unaffected by the first interview. Something might be done to test the consistency of psychiatric observers by presenting two psychiatrists with the same detailed clinical record or by allowing each to listen to a tape recording or a sound film of the other's interviews. Until such developments become commonplace, however, it may be simplest to rest content with the information gained from the prevalence estimates obtained from random samples of the same population examined by different observers, either as individuals or classified according to background and training.

Standardization of diagnostic techniques

All these sources of variation can be investigated in any situation where the clinical material, human or recorded, can be presented in turn to different observers. From each age- and sex-group of the population to be surveyed is drawn a sample of an equal number of individuals (or the corresponding clinical material). These are seen in turn by equal numbers of different observers of each type, e.g., psychiatrist and lay. In this way, the mean frequency of positive findings, e.g., on personal or family history, made by different types of observer is based on samples balanced in respect of age and sex. Comparisons can then be readily made between the results obtained using observers of different types. Similarly, interaction between observer and observed can be investigated as before, e.g., by comparing the assessment made of clinical material relating to the same set of male subjects by male and female lay investigators.

A detailed review of each item in the records used in such a study will often reveal the nature of the disagreement between either individuals or types of observer. Differing emphasis on items dealing with family history, for example, may explain a divergent assessment of emotional stability made by psychiatrists and lay social workers. Once these sources of discrepancy have been uncovered, their effect can be minimized.

In the standardization of diagnosis, agreement between observers on a detailed specification of each diagnostic category is an obvious first step. Such specification can be powerfully aided by examples of clinical material typical of each category—a sort of “standard case” against which others found in the survey can be classified. In a questionnaire, the drafting of each item needs meticulous care. All the possible alternative responses should be foreseen and classified in the training instructions. Finally, the ultimate categorization may be made to depend on independent tests or assessments. In the milder degrees of schizophrenia, for example, psychological tests of intelligence or records of school attainment may be decisive. In less well-defined conditions, review of the clinical protocols by an independent consultant, ignorant of the views of the other members of the team, will ensure some uniformity of categorization in borderline cases. Whatever the trouble involved, experience in other fields of epidemiological work strongly affirms the need for such standardization in technique and the specification, in reports of survey work done, of the precise diagnostic criteria used. Dohrenwend (1957) has given an account of the construction and testing of the screening procedure used in the Stirling County study. The results of preliminary testing by a psychosomatic inventory are compared with findings from

hospital records and information given by local doctors. Essen-Möller (1956) in his survey of an area in southern Sweden, correlated complaints of physical disorder with personality ratings and clinical assessments of intellectual capacity with marks in school examinations.

Some simple statistical points in the interpretation of results

Whether he conducts such inquiries himself or not, the psychiatrist interested in epidemiological studies should be able to assess the evidence presented in the reports of others. Certainly, sophisticated design and analysis of data usually mean an appeal to the opinion of the professional statistician. But most of the major errors in the interpretation of results come, not from the inappropriateness of the particular statistical techniques used, but from failure to appreciate the simple and more subtle sources of bias. These are often more obvious to the psychiatrist than to a statistician who may be unfamiliar with the clinical background. Some of the points already made thus bear re-emphasis. There are four basic questions which are always worth asking:

1. Is there an appropriate measure of relative risks? Quite apart from the elementary error of failing to relate the number of cases to the population at risk, the computation of that exposure to risk is fundamental. Knowledge of the natural history of the particular disease may suggest that only one age- and sex-group is effectively exposed. Again, the same care is needed to define cases and populations exposed and the duration of effective exposure in the two populations being compared.

2. Is there any bias in the sampling procedure used? Where any deviation from strictly random sampling is involved, special care is needed to avoid the influence of some unforeseen source of bias.

3. Are cases and controls essentially alike in all characteristics likely to be relevant except those under review? This question is especially appropriate where other hospital patients have been used as "controls". Retrospective inquiries will give valid answers only in so far as the cases studied are representative of all such cases arising in a defined population and as the controls are typical of that same population.

4. Are any differences observed likely to be due to chance? This is primarily a technical question answerable by technical tests of significance. These test answer that question and no other. They give no proof of the alternative hypothesis proposed, nor do they necessarily imply that any differences which are large enough to be unlikely to be fortuitous are therefore large enough to be of practical importance.

EPIDEMIOLOGICAL EXPERIMENTS IN DISEASE CONTROL

For many types of mental disorder, prevention or control is still a prospect rather than an achievement. Gruenberg (1957), however, has emphasized the essentially preventable nature of neuropsychiatric disabilities linked with physical conditions, like pellagra or lead poisoning, whose prevention is already part of public health work. Intelligent assessment of the degree of success of any systematic attempt to control such mental complications of brain injury or poisoning could well follow conventional lines.

Supervision of mental disorder control

Death registration can play only a minor part in the observation of time-trends in mental disorder incidence in relation to changes in control methods. Morbidity records, based on regular mandatory notification, are the traditional index in infectious disease control; and their potential and limitations would apply equally in the field of mental health. On the basis of current trends in the incidence of disease under what we believe to be stable conditions, we estimate the "expectancy" for the number of cases in each successive time period and then define the likely limits of variation round that estimate. A convenient rule of thumb for a steady endemic incidence of a few cases gives "control limits" of \pm twice the square root of the expected number. Thus, if the number of cases of pellagra psychosis reported in recent years was at a steady level, averaging 16, the appropriate limits of chance variation would be $16 \pm 2 \times \sqrt{16} = 8$ to 24. If later observation falls outside these limits of chance fluctuation, the epidemic situation has presumably changed for better or for worse. When improvement coincides with preventive action, we may consider whether we are observing cause and effect.

Unfortunately, it is often impossible to be certain that these changes are not more apparent than real. The ecological equilibrium of disease may be upset by simultaneous social or industrial improvement. Con-

versely, the promotion of the preventive itself may, by engendering interest in the disease concerned, inflate the reporting rate.

For these reasons, the crucial test of preventive action usually has to come from the deliberate experiment. The principles of conduct in such field trials have been frequently set out, e.g., Bradford Hill (1958), but they may be usefully summarized with special reference to mental health. Here, four kinds of approach in "disease control" are likely to arise—the detection of susceptibles, disease modification by education, anticipation by supervision, and the changing of environment. In industry, selection procedures based on tests of intelligence and emotional stability may be introduced to reject applicants for employment in jobs where stability under stress is of particular importance. Beyond the selection stage, training schemes and a process of gradual habituation to harassing conditions may be used to prevent the acute neurosis. Dietary supplements, such as reinforced bread or milk, can mitigate the effects on the central nervous system of either alcohol or lead. Where a toxic hazard like lead may cause mental disorder, the physical methods used in its control, such as the search for punctate basophils in regularly taken blood samples, will be equally helpful in anticipating the toxic effects on the central nervous system. In the same context, attempts to reduce atmospheric concentrations of lead or other central nervous system poisons in industry will be tantamount to an environmental control of encephalopathy. Some of the problems involved in the field-testing of these alternative control methods are common to all. The first has already been indicated; comparisons in time, before and after the introduction of some preventive action, are notoriously treacherous. Unrelated contemporary events—such as a change in the economic climate altering the quality of recruits to industry—may affect the incidence of mental disorder in an occupational group quite independently of a newly-introduced psychiatric selection procedure. Comparisons involving volunteers are another source of error. The men who elect to take dietary supplements like milk to prevent plumbism are likely to differ in mental constitution from the less responsive remainder. To use the latter's mental disorder experience as a "control" on the effects of dietary supplements would clearly result in misinterpretation.

Controlled field trials

The most certain method of assessment is the carefully planned and controlled field trial. Often novel methods, e.g., of child care and training, are widely introduced without an adequate comparison with either the previously favoured policy or, indeed, with the practice of "masterly inactivity". Once such new methods become the established practice, the opportunity for their strict assessment has passed, for the withholding

of such methods in randomly selected subjects could not be generally acceptable. Comparative field trials should therefore be an essential preamble to any important change in preventive policy.

The principles of experimentation apply equally to series of individuals and to series of homogeneous social or other groups. It may be impossible to draw samples of either from the whole population, and volunteers who understand the implications of the procedures involved will have to be used. To overcome this inherent bias due to volunteering, these volunteers must be allocated randomly to the two methods of prophylaxis being compared, so that both series are alike in volunteering and differ in other respects by no more than is likely to be due to chance. In other words, random allocation within this volunteer group will ensure that there is no consistent bias entering into the subsequent comparison of experience in the two series. Greater equality between the series can be achieved by dividing the original volunteers into strata which are likely, on *a priori* reasoning, to differ appreciably in their mental disease incidence. Within, for example, each age- and sex-stratum, the volunteers are then divided randomly into two series. This ensures that, at least in these material respects, the two groups are completely alike. Whatever the method of allocation used, it is always worth while checking that the two series thus produced are alike, particularly in those factors which have not been taken into account by stratification by age or sex but which are likely to be material in the outcome of the trial.

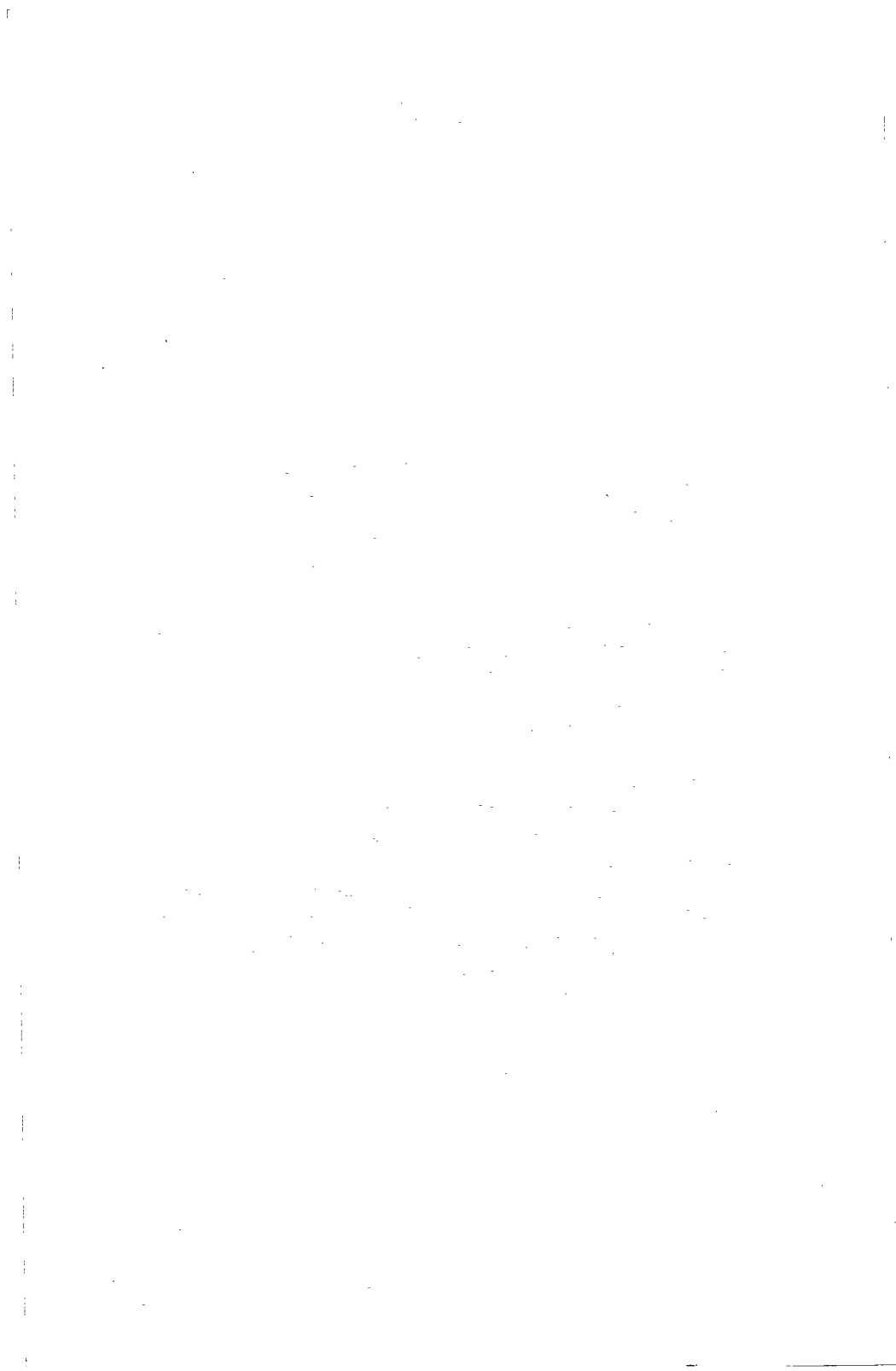
In preventive psychiatry, it may not be practicable to apply apparently identical procedures, either to two individuals or to two social or other groups. To this extent, field studies in psychiatry differ from trials of vaccines, where two types are being compared and the subject need not know the precise nature of the preparation inoculated. Nevertheless, although no exact parallel with the placebo of therapeutic trials may be possible, the rigorous execution of all the other aspects of the trial will minimize the importance of serious bias from this source.

During the period of observation and in the final assessment of result, three equalities are essential—reasonable equality of exposure in the two series to environmental and other stresses, equality in the completeness of follow up and ascertainment of disorder, and equality in the methods of clinical assessment. In a study of maternal nutrition during pregnancy in relation to mental deficiency, behaviour disorders, and epilepsy in the offspring, for example, the children of mothers *not* given extra food would be followed up by regular visits just as frequently as those of mothers in the experimental group whose diet had been supplemented. Equality of exposure to environmental stresses might be measured by recording with equal care both the frequency of serious or trivial

accidents and intercurrent illness and death in siblings or parents. Similarly, in the final assessment of adjustment and intelligence, assessment should be made in both series by the same psychiatrist and psychologist, who should each give his verdict without knowledge either of the other's assessments or of the series to which the child belonged. By such methods, conclusions can be firmly based and the whole practice of disease control or prevention in psychiatry placed on a sound foundation.

POSTSCRIPT

In retrospect, the tone of this commentary on past experience in using epidemiological methods for the study of mental disorder may seem critical—at times to the point of nihilism. This was certainly not the intention behind its production. The pioneer studies described have made a major contribution to our understanding of the observed patterns of mental disorder in human populations. Their imperfections were usually as obvious to the authors as to the armchair critic. But if this relatively novel application of epidemiological method is to be usefully productive, its underlying assumptions and practical limitations have to be recognized. Once they are appreciated, much can be done to minimize the handicaps inevitable in an observational science. Absolute and final proof may be an unattainable ideal; but the value of the comparison of the disease experience in different conditions may lie mainly in the discrimination between unlikely and likely causes of mental disorder. The findings may expose the falsity of beliefs about, perhaps, the relative importance of environmental and innate sources of disease, which determine some current methods of prevention or treatment. In a rapidly changing world, there are countless opportunities for the detection of environmental and personal factors in disease causation which may be susceptible to change or removal. By a rigorous development of investigative technique, epidemiological studies can at least define the limits of our knowledge and then build up a foundation of fact and sound theory for rational policies in the control and cure of mental disorder.



REFERENCES

- Adelstein, A. M. (1952) *J. R. stat. Soc.*, **115**, 354
 Arbous, A. G. & Kerrich, J. E. (1951) *Biometrics*, **7**, 340
 Berkson, J. (1946) *Biomet. Bull.*, **2**, 47
 Birkelo, C. C. et al. (1947) *J. Amer. med. Ass.*, **133**, 359,
 Boeters, D. (1936) *Z. Neurol.*, **155**, 675
 Böök, J. A. (1953) *Acta genet. (Basel)*, **4**, 1
 Böök, J. A. (1955) *Eugen. Quart.*, **2**, 174
 Bremer, J. (1951) *Acta psychiat. (Kbh.) Suppl.* 62, p. 1
 Brugger, C. (1933) *Z. Neurol.*, **145**, 516
 Buck, C. (1955) *Canad. Serv. med. J.*, **11**, 791
 Carstairs, G. M. & Brown, G. W. (1958) *J. ment. Sci.*, **104**, 72
 Clausen, J. A. & Kohn, M. L. (1954) *Amer. J. Sociol.*, **60**, 140
 Cochrane, W. G. (1951) *Amer. J. publ. Hlth*, **41**, 6
 Dahlberg, G. & Stenberg, S. (1931) *Z. Neurol.*, **133**, 447
 Dohrenwend, B. P. (1957) *Amer. Psychol.*, **12**, 2
 Dorn, H. F. (1957) *Publ. Hlth Rep. (Wash.)*, **72**, 1043
 Douglas J. W. B. & Blomfield, J. M. (1958) *Children under five*, London
 Downes, J. & Simon, K. (1954). *Milbank mem. Fd Quart.*, **32**, 1
 DUNBAR, H. F. (1946) *Emotions and bodily changes*, London
 Elkind, H. B. & Doering, C. R. In: Boston Psychopathic Hospital (1925-34) *Schizophrenia, Studies from the Boston Psychopathic Hospital*, p. 75
 Essen-Möller, E. (1956) *Acta psychiat. scand.*, 100 Suppl.
 Faris, R. E. L. & Dunham, N. W. (1939) *Mental disorders in urban areas*, Chicago
 Felix, R. H. & Kramer, M. (1952) *Publ. Hlth Rep. (Wash.)*, **67**, 152
 Felix, R. H. & Kramer, M. (1953) *Ann. Amer. Acad. pol. soc. Sci.*, 286, 5
 Fisher, R. A. & Yates, F. (1938) *Statistical tables for biological, agricultural and medical research*, London
 Fraser, R. et al. (1947) *Rep. industr. Hlth Res. Bd (Lond.)*, **90**
 Fremming, K. H. (1947) *Morbid risk of mental diseases in an average Danish population*, Copenhagen (English Edition, 1951)
 Fry, J. (1957) *Brit. med. J.*, **2**, 1453
 General Register Office, England and Wales (1954) *Measurement of morbidity* (Studies on Medical and Population Subjects, No. 8), London
 Goldberger, J. (1916) *J. Amer. med. Ass.*, **66**, 471
 Goldhamer, H. & Marshall, A. (1953) *Psychosis and civilization*, Glencoe, Ill.
 Great Britain. Air Ministry (1946). *Psychological disorders in flying personnel of the Royal Air Force investigated during the war, 1939-1945*, London (Air Publication No. 3139)

- Greenwood, M. et al. (1936) *Spec. Rep. Ser. med. Res. Coun. (Lond.)*, **209**
- Greenwood, M. & Yule, G. U. (1920) *J. Roy. stat. Soc.*, **83**, 255
- Gruenberg, E. M. (1957) *Amer. J. publ. Hlth*, **47**, 944
- Haeser, H. (1882) *Lehrbuch der Geschichte der Medicin*, Jena
- Halliday, J. (1948) *Psychosocial medicine*, London
- Heasman, M. A. et al. (1958) *Brit. J. industr. Med.*, **15**, 141
- Hecker, J. F. C. (1832) *Die Tanzwuth, eine Volkskrankheit im Mittelalter*, Berlin
- Hill, A. B. (1955) *Principles of medical statistics*, London
- Hill, A. B. (1958) *J. R. Inst. publ. Hlth*, **21**, 177
- Hollingshead, A. B. & Redlich, F. C. (1958) *Mental disease and social class*, New York
- Hyde, R. W. & Kingsley, L. V. (1944) *New Engl. J. Med.*, **231**, 543
- Jennings, D. (1940) *Lancet*, **1**, 395, 444
- Jost, H. see Diem, O. (1905) *Arch. Rassenbiol.*, **2**, 215
- Kallmann, F. J. (1946) *Amer. J. Psychiat.*, **102**, 522
- Kallmann, F. J. (1950) *Amer. J. Psychiat.*, **106**, 501
- Kallmann, F. J. (1953) *Amer. J. Psychiat.*, **109**, 491
- Kendall, M. G. & Smith, B. B. (1939) *Tables of random sampling numbers*, London (Tracts for Computers No. 24)
- Kilpatrick, S. J. et al. (1955) *Ulster med. J.*, **24**, 113
- Klemperer, J. (1933) *Z. Neurol.*, **146**, 277
- Koller, J. (1895) *Arch. Psychiat. Nervenkr.*, **27**, 268
- Ledermann, S. (1953) *Concours méd.*, **75**, 1485; 1583; 1675; 1767
- Lemkau, P. et al. (1941) *Ment. Hyg. (N.Y.)*, **25**, 624
- Lilienfeld, A. M. & Pasamanick, B. (1954) *J. Amer. med. Ass.*, **155**, 719
- Lin, T. (1953) *Psychiatry*, **16**, 313
- Logan, W. P. D. (1956) *Proc. roy. Soc. Med.*, **49**, 495
- London Transport Executive (1956) *Health in industry*, London
- Luxemburger, H. (1928) *Z. Neurol.*, **112**, 331
- McIntosh, R. et al. (1954) *Pediatrics*, **14**, 505
- MacMahon, B. & McKeown, T. (1952) *Brit. J. soc. Med.*, **6**, 265
- Milbank Memorial Fund (1950) *Epidemiology of mental disorder*, New York. (Papers presented at a round table at the 1949 annual conference of the Milbank Memorial Fund, November 16-17, 1949)
- Moser, C. A. (1958) *Survey methods in social investigation*, London
- Ødegaard, O. (1952) *Psychiat. Quart.*, **26**, 212
- Penrose, L. S. (1939) *Brit. J. med. Psychol.*, **18**, 1
- Penrose, L. S. (1958) *J. ment. Defic. Res.*, **1**, 4
- Registrar General, England and Wales (1958) *The Registrar General's decennial supplement. Occupational mortality*, London
- Reid, D. D. (1948) *Brit. J. soc. Med.*, **2**, 123
- Roberts, J. A. F. (1948) *Ann. Eugenics*, **14**, 109
- Robinson, W. D. (1950) *Amer. sociol. Rev.*, **15**, 351
- Ryle, J. & Russell, W. T. (1949) *Brit. Heart J.*, **11**, 370
- Sainsbury, P. (1955) *Suicide in London*, London (Maudsley Monographs No. 1)
- Schulz, B. (1936) *Methodik der medizinischen Erforschung*, Berlin
- Scottish Council for Research in Education (1949) *The trend of Scottish intelligence*, London
- Slater, E. (1953) *Spec. Rep. Ser. med. Res. Coun. (Lond.)*, **278**
- Smiley et al. (1955) *Milbank mem. Fd Quart.*, **33**, 213
- Smith, M. & Leiper, M. (1936) *Rep. industr. Hlth Res. Bd (Lond.)*, **75**, Part 1.
- Srole, L. & Langner, T. S. (1957) Paper read to American Sociological Society, Washington

- Strömngren, E. (1935) *Z. Neurol.*, **153**, 784
Strömngren, E. (1938) *Acta psychiat. (Kbh.)*, Suppl. 19
Strömngren, E. (1950) In: International Congress of Psychiatry, Paris, 1950. *Comptes rendus des séances*, Paris
Symonds, C. P. (1943) *Brit. med. J.*, **2**, 703, 740
Tippett, L. H. C. (1927) *Random sampling numbers*, London (Tracts for Computers No. 15)
Vermeulen, G. (1943) *J. belge Neurol. Psychiat.*, **5** & **6**
Weinberg, W. (1925) *Meth. Tech. Stat.*, **1**, 138
Winston, E. B. (1931) *Ment. Hyg. (N.Y.)*, **15**, 761
Wolf, G. (1928) *Z. Neurol.*, **117**, 728
Yule, G. U. & Kendall, M. G. (1950) *Introduction to the theory of statistics*, London,
-

