WORLD PSYCHIATRIC ASSOCIATION
ROYAL MEDICO-PSYCHOLOGICAL ASSOCIATION

PSYCHIATRIC EPIDEMIOLOGY

Proceedings of the International Symposium held at Aberdeen University 22–5 July 1969

EDITED BY
E. H. HARE AND J. K. WING

PUBLISHED FOR THE
NUFFIELD PROVINCIAL HOSPITALS TRUST
BY THE OXFORD UNIVERSITY PRESS
London New York Toronto
1970

Oxford University Press, Ely House, London W1
GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
CAPE TOWN SALISBURY IBADAN NAIROBI DAR ES SALAAM LUSAKA ADDIS ABABA
BOMBAY CALCUTTA MADRAS KARACHI LAHORE DACCA
KUALA LUMPUR SINGAPORE HONG KONG TOKYO

SBN 19 721360 x

Designed by Bernard Crossland

PRINTED IN GREAT BRITAIN
BY HAZELL WATSON AND VINEY LTD
AYLESBURY, BUCKS

PREFACE

The Symposium on Psychiatric Epidemiology, held at Aberdeen University in July 1969, was jointly sponsored by the Royal Medico-Psychological Association and the World Psychiatric Association. A small organizing committee, consisting of Professor Ernest Gruenberg, Dr E. H. Hare, Dr George Innes, Professor Kenneth Rawnsley, and Dr John Wing, was set up to represent the two associations. There has been no equivalent meeting since the Milbank Memorial Fund held a Round Table on Causes of Mental Disorders in 1959, and the Nuffield Provincial Hospitals Trust Symposium in 1961.

During the past ten years, numerous epidemiological studies have been undertaken, both of aetiology and of treatment, but the object of the symposium was not so much to review this accumulation of knowledge as to present the current work of investigators in the field, and to discuss certain problems of method which concern most epidemiologists. Inevitably, several investigators who were invited could not attend the meeting. There was a further restriction owing to the fact that certain topics, such as delinquency, drug dependency, sexual deviations, personality disorders, and clinical syndromes found particularly in certain subcultures, were not included, although they have all been investigated using epidemiological methods. The term 'epidemiology' itself was not defined in advance but there was, in fact, a fair uniformity of usage. Professor Mechanic points out that it is not uncommon for social scientists 'to discover, on being introduced to the field of epidemiology, that this is what they have been doing all their professional lives'. One of the chief values of a strong discipline of epidemiology within psychiatry, which the symposium was designed to enhance and exploit, is that it provides standards against which to measure scientific performance. If there is an informed and critical audience, inexactly formulated hypotheses, inappropriate methods of investigation, and over-enthusiastic generalizations from results, will become less prevalent. The main purpose of the symposium was to provide an opportunity for scientists working in the field to meet and discuss current work in the hope that their future work and intercommunication would be stimulated and enriched.

It was thought that a substantial part of the proceedings of the symposium merited publication even though several papers had to be cut considerably in length. The presentations by Dr Dohrenwend and Dr Hagnell were particularly affected but the full papers are available to interested workers on application to the authors. Another problem was whether to include an account of discussion from the floor. The editors finally decided against this since it seemed reasonable to allow authors to modify their papers for publication and it would have been very difficult, in any case, to have reduced the tape-recorded discussions to a readable length. As it is, this volume, from Professor Mechanic's broad survey to Dr Barraclough's brief but pithy report, bears witness to the present healthy state of psychiatric epidemiology.

REFERENCES

GRUENBERG, E. M., and HUXLEY, M. (1961). *Causes of Mental Disorder: A Review of Epidemiological Knowledge* (New York: Milbank Memorial Fund).

HILL, D. *et al.* (1962). *The Burden on the Community and the Epidemiology of Mental Illness* (London: Oxford University Press).

ACKNOWLEDGEMENT

It is a pleasure to acknowledge a grant of £250 from the Rowntree Charitable Trust to assist with the expenses of a number of participants.

CONTENTS

LIST OF PARTICIPANTS

DR A. M. ADELSTEIN, MD, DPH, General Register Office, Somerset House, London, WC2.

PROFESSOR Y. ALANEN, MD, Department of Psychiatry, University of Turku, Turku 6, Finland.

DR JACQUELINE DE ALARCON, PH D, MRC Unit, Graylingwell Hospital, Chichester, Sussex.

DR RICHARD DE ALARCON, MD, DPM, MRC Unit, Graylingwell Hospital, Chichester, Sussex.

PROFESSOR E. MAURICE BACKETT, FRCP, Department of Social Medicine, University Medical Buildings, Foresterhill, Aberdeen.

DR ANITA BAHN, SC D, Woman's Medical College, 3300 Henry Avenue, Philadelphia, Pa. 19129, U.S.A.

DR J. A. BALDWIN, MD, Oxford Record Linkage Study and Unit of Clinical Epidemiology, Oxford Regional Hospital Board, Old Road, Headington, Oxford.

DR B. M. BARRACLOUGH, MB, MRACP, DPM, 7c Parchment Street, Chichester, Sussex.

DR JAMES BARRETT, MD, Institute of Psychiatry, De Crespigny Park, London, SE5.

MRS MILDRED B. BECK, MSW, Research Support Section, Center for Epidemiologic Studies, National Institute of Mental Health, 5454 Wisconsin Avenue, Wt. 304, Chevy Chase, Md. 20015, U.S.A.

DR DOUGLAS BENNETT, MD, DPM, The Maudsley Hospital, Denmark Hill, London, SE5.

DR J. L. T. BIRLEY, MRCP, DPM, Social Psychiatry Research Unit, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR ARTHUR BOWEN, DM, Bootham Park, York.

DR E. R. BRANSBY, Department of Health and Social Security, 14 Russell Square, London WC1.

MISS EILEEN M. BROOKE, M SC, DPA, 10 rue Grenus, Geneva, Switzerland.

DR GEORGE W. BROWN, PH D, Bedford College Annexe, Peto Place, Marylebone Road, London, NW1.

PROFESSOR G. M. CARSTAIRS, MD, FRCPE, Department of Psychological Medicine, Morningside Park, Edinburgh 10.

DR A. K. CLARKE, MB, BS, 26 Mount Terrace, Stepney, London, E1.

DR BRIAN COOPER, MD, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR JOHN E. COOPER, BM, MRCP, DPM, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR J. R. M. COPELAND, MB, MRCP, DPM, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR M. VON CRANACH, MD, Institute of Psychiatry, Denmark Hill, London, SE5.

DR BARBARA S. DOHRENWEND, PH D, 1270 Fifth Avenue, New York, NY 10029, U.S.A.

DR BRUCE P. DOHRENWEND, PH D, 1270 Fifth Avenue, New York, NY 10029, U.S.A.

DR J. DONNELLY, MB, Scottish Home and Health Department, 79 Lauriston Place, Edinburgh.

DR J. W. B. DOUGLAS, BM, MRC Unit, London School of Economics, Houghton Street, London, WC2.

PROFESSOR H. W. DUNHAM, PH D, Wayne State University, College of Liberal Arts, Detroit, Mich. 48202, U.S.A.

DR M. R. EASTWOOD, MB, DPM, Institute of Psychiatry, De Crespigny Park, London, SE5.

PROFESSOR E. ESSEN-MÖLLER, MD, Alsback, Lysekil 45300, Sweden.

PROFESSOR LUIGI FRIGHI, Via Baccio Pontelli, 16, 00153 Rome, Italy.

DR THOMAS FRYERS, MB, DPH, Department of Social and Preventive Medicine, Manchester University, Manchester.

DR ELMER A. GARDNER, MD, Temple University Health Science Centre, Philadelphia, Pa., U.S.A.

DR DAVID P. GOLDBERG, Institute of Psychiatry, De Crispign

DR DAVID P. GOLDBERG, MRCP, DPM, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR E. M. GRUENBERG, MD, Psychiatric Epidemiology Unit, Hudson River State Hospital, Poughkeepsie, NY 12601, U.S.A.

PROFESSOR H. HÄFNER, Voss Strasse, 4, 69 Heidelberg 1, West Germany.

DR O. HAGNELL, MD, Department of Psychiatry, University of Lund, St. Lars Sjukhus, Lund, Sweden.

DR VIRGINIA HANNON, SC D, Epidemiology Laboratory, Research Center, Rockland State Hospital, Orangeburg, New York, NY 10962, U.S.A.

DR E. H. HARE, MD, The Maudsley Hospital, Denmark Hill, London, SE5.

DR M. HARPER, MB, MANZCP, Department of Psychological Medicine, University of Queensland, Australia.

PROFESSOR A. SCOTT HENDERSON, MD, MRCP, DPM, Department of Psychiatry, Melrose, 141 Howpden Road, Hobart, Tasmania, 7000.

PROFESSOR J. HOENIG, MD, MRCP, DPM, Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

PROFESSOR RAYMOND ILLSLEY, PH D, Department of Sociology, University of Aberdeen, Centre for Social Studies, Westburn Road, Aberdeen.

DR GEORGE INNES, MD, DPH, Department of Social Medicine, University Medical Buildings, Foresterhill, Aberdeen.

DR ASSEN JABLENSKY, MD, The Maudsley Hospital, Denmark Hill, London, SE5.

DR N. JUEL-NIELSEN, Department of Psychiatry, O.A.B.S., Odense, Denmark.

DR MARTIN M. KATZ, PH D, Clinical Research Branch, National Institute of Mental Health, 5454 Wisconsin Avenue, Chevy Chase, Md. 20015, U.S.A.

DR PREDRAG KALICANIN, Institute for Mental Health, Zavod Za Mentalno Zdravlie, Palmoticeva 37, Belgrade, Yugoslavia.

DR H. KATCHADOURIAN, MD, Department of Psychiatry, Stanford Medical Centre, Stanford, Calif. 94305, U.S.A.

PROFESSOR W. I. N. KESSEL, MD, FRCP, Department of Psychiatry, University of Manchester, Swinton Grove, Manchester 13.

PROFESSOR L. G. KILOH, MD, FRCP, School of Psychiatry, University of New South Wales, Australia.

PROFESSOR D. DAVID KIRK, PH D, Professor of Sociology, University of Waterloo, Montreal, Canada.

DR RUTH V. KIRK, M SC, Sociological Laboratory, University of Waterloo, Montreal, Canada.

PROFESSOR R. J. KLEINER, PH D, Temple University, College of Liberal Arts, Philadelphia, Pa. 19122, U.S.A.

DR N. KREITMAN, MD, DPM, MRC Unit, Department of Psychiatry, Royal Edinburgh Hospital, Edinburgh.

DR EINAR KRINGLEN, MD, Psykiatrisk Klinikk, Haukeland Sykehus, 5000 Bergen, Norway.

DR J. P. LEFF, MB, MRCP, DPM, Social Psychiatry Research Unit, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR VILLE LEHTINEN, Department of Psychiatry, University of Turku, Turku 6, Finland.

DR KATHERINE LEVY, MB, Medical Research Council, 20 Park Crescent, London, W1.

PROFESSOR SIR AUBREY LEWIS, MD, FRCP, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR R. JAMES MCCORKEL, MA, Apt. 9-J, 511 Chesterfield, Nashville, Tenn. 37212, U.S.A.

PROFESSOR DAVID MECHANIC, PH D, Department of Sociology, University of Wisconsin, Madison, Wis. 53706, U.S.A.

PROFESSOR W. M. MILLAR, MD, FRCPE, Department of Mental Health, University Medical Buildings, Foresterhill, Aberdeen.

PROFESSOR J. N. MORRIS, D SC, FRCP, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1.

DR H. B. M. MURPHY, MD, PH D, McGill University, Beatty Hall, Montreal, 2, P.Q., Canada.

PROFESSOR J. S. NEKI, MB, DPM, Department of Psychiatry, All-India Institute of Medical Sciences, New Delhi, India.

DR J. R. NEWBROUGH, PH D, Department of Psychology, Peabody College, Nashville, Tenn. 37203, U.S.A.

PROFESSOR Ø. ØDEGAARD, MD, Gaustad Hospital, Vinderen, Oslo 3, Norway.

DR P. C. OLLEY, MB, Department of Mental Health, University Medical Buildings, Foresterhill, Aberdeen.

DR JOHN ORLEY, BM, 717 Chelsea Cloisters, Sloane Avenue, London, SW3.

PROFESSOR DESMOND A. POND, MD, FRCP, The London Hospital Medical College, Whitechapel, London, E1.

DR NICOLE QUEMADA, Inserm, Section Psychiatrie, 15 rue de Abondances, 92 Boulogne, France.

DR WILTRUD RANANAUER-HAFNER, Voss Strasse 4, 69 Heidelberg 1, West Germany.

PROFESSOR K. RAWNSLEY, MB, FRCP, Department of Psychological Medicine, Whitchurch Hospital, Cardiff.

DR T. REICH, MD, Department of Psychiatry, Barnes and Renard Hospitals, 4940 Audubon Avenue, St. Louis, Mo. 63110, U.S.A.

PROFESSOR D. D. REID, MD, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1.

FRAU DR HELGA REIMANN, Social Psychiatrische Klinik der Universitaet, 69 Heidelberg 1, Voss Strasse 4, West Germany.

DR A. RICHMAN, MD, College of Physicians and Surgeons, 722 West 168th Street, New York, NY 10032, U.S.A.

PROFESSOR LEE N. ROBINS, PH D, Department of Psychiatry, Washington University Medical School, 4940 Audubon Avenue, St. Louis, Mo. 63110, U.S.A.

PROFESSOR MARTIN ROTH, MD, FRCP, Department of Psychological Medicine, University of Newcastle upon Tyne, Newcastle upon Tyne.

DR M. L. RUTTER, MD, MRCP, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR R. SADOUN, Institute Nationale de la Santé et de la Recherche Medicale, 3 rue Leon-Bonnat, Paris 16e, France.

DR PETER SAINSBURY, MD, MRCP, MRC Clinical Psychiatry Research Unit, Graylingwell Hospital, Chichester, Sussex.

DR NORMAN SARTORIUS, MD, PH D, Mental Health Unit, World Health Organization, Geneva, Switzerland.

DR LAWRENCE SHARPE, MB, DPM, New York Psychiatric Institute, 722 West 168th Street, New York, NY 10032, U.S.A.

DR C. SKODA, MD, Psychiatric Research Institute, Prague 8, Bohnice, Czechoslovakia.

PROFESSOR MICHAEL SHEPHERD, DM, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR ZENA STEIN, MB, School of Public Health and Administrative Medicine, 600 West 168th Street, New York, NY 10032, U.S.A.

DR MERVYN W. SUSSER, MB, School of Public Health and Administrative Medicine, 600 West 168th Street, New York, NY 10032, U.S.A.

DR PEKKA TIENARI, MD, Koskelantie 106, Oulu 52, Finland.

DR R. JAY TURNER, PH D, Temple Community Mental Health Center 3501 North 16th Street, Philadelphia, Penn. 19140, U.S.A.

DR ERKKI VAISANEN, Oulu University Central Hospitals, Oulu, Finland.

DR VANA, MD, Psychiatric Research Institute, Prague 8, Czechoslovakia.

DR JOHN K. WING, MD, PH D, MRC Social Psychiatry Research Unit, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR LORNA WING, MD, MRC Social Psychiatry Research Unit, Institute of Psychiatry, De Crespigny Park, London, SE5.

DR D. VON ZERSSEN, MD, Max-Planck Institute fur Psychiatrie, 8 München 23, Kraepelinstrasse 10, Germany.

I

PROBLEMS AND PROSPECTS IN PSYCHIATRIC EPIDEMIOLOGY

Problems and prospects in psychiatric epidemiology[1]

DAVID MECHANIC
Professor and Chairman, Department of Sociology,
University of Wisconsin, U.S.A.

The concern of epidemiology is the distribution of disease, defects, and disabilities in populations, and the various personal and environmental factors that affect how such conditions become manifest. Since the organization of services and other forms of intervention in countering morbidity affect the manner in which disease occurs and is distributed, it is not uncommon for the epidemiologist to study as well intervention systems and the impact of providing services on the population. Epidemiology is distinguished from many other research fields studying similar problems and using similar techniques of inquiry by the fact that it focuses on a disease, defect, or disability as its dependent variable. Thus epidemiology has a strong applied emphasis and is pursued mainly to elicit clues as to the aetiology of disease or to plan meaningful health services. The two most commonly used measures in epidemiology—incidence and prevalence—pertain specifically to each of these two goals.

In short, the epidemiologist is not differentiated from the demographer, human ecologist, or survey researcher by a unique set of methods. It is not uncommon for sociologists, ecologists, and social psychologists to discover, on being introduced to the method of epidemiology, that this is what they have been doing all of their professional lives. What distinguishes him is that the epidemiologist focuses on disease rather than on other dependent variables such as migration, social mobility, urbanization, segregation, and the like. And it is this emphasis on disease as the dependent variable—the core of

---

[1] Various sections of this paper are taken from David Mechanic, *Mental Health and Social Policy* (Englewood Cliffs: Prentice-Hall, 1969), where they are developed in greater detail. This paper was written, in part, with the assistance of Grant R11 MH–02014 from the National Institute for Mental Health.

epidemiological effort—that betrays the weakness of the current state of epidemiological efforts in the psychiatric area.

It is a cardinal principle of epidemiology that investigations are only as good as the techniques used in case-identification. Implicit in epidemiological investigation is the assumption that the dependent variable studied can be reliably differentiated from related variables, and that measurement is sufficiently refined so that the condition under inquiry can be graded in terms of severity, disability, or some other indicator of the degree of impact of the problem. Like a ship that begins on the wrong course, an epidemiological investigation that uses weak and unreliable indicators of its dependent variable is unlikely to reach its desired destination. Moreover, reliability is not enough; adequate case-finding techniques must be consistent with the accepted conceptions of pathology and disability among professional workers. As much as some investigators may find it convenient to study hypercholesterolemia or high blood pressure instead of ischemic heart disease, or stomach acidity instead of stomach ulcers, we must not confuse one with the other. The factors correlated with hypercholesterolemia and stomach acidity are not the identical factors associated with the occurrence of ischemic heart disease and stomach ulcers, and substituting one for the other, however convenient, may be misleading. In the case of blood cholesterol level, for example, the epidemiological evidence itself, as Yerushalmy (36) has noted, refutes the validity of substituting one condition for the other:

> It is easy to demonstrate that the blood cholesterol level cannot be accepted as a substitute for coronary heart disease. The two are not synonymous and interchangeable. For example, the blood cholesterol level does not continue to increase with age in men past the age of 50, while the death rate from coronary heart disease continues to increase. The blood cholesterol levels are almost identical for men and women, while the coronary heart disease rates are much different for the two sexes. There is a lack of correlation between serum cholesterol levels and the degree of atherosclerotic process found at autopsy of persons dying from noncoronary heart disease . . . Perhaps the strongest argument against the 'substitution game' is that reduction of the serum cholesterol level does not seem to influence the risk of coronary disease (36).

Similarly, when one investigates more global categories of illness and disability, it is essential that the condition itself be stratified in terms of stages of disease and levels of seriousness. Different types of mental illness are the result of varying causal agents, and in planning health services to meet a problem it is necessary to distinguish such conditions in terms of what we know about successful intervention and varying levels of need for services. It would be ludicrous if, in measuring population needs, those planning comprehensive medical clinics included the common cold in their definition of morbidity. Yet it is commonplace in psychiatric epidemiology to treat such varying problems as chronic psychoses, incapacitating neurotic conditions, life problems, and psychophysiological symptoms. In determining the needs of the population for mental health services, for example, the use of an all-inclusive concept of mental illness fails to separate problems that necessitate immediate care from those that are mild and transitory and it does not allow for the different aetiology and character of the mental disorders.

In viewing various results from epidemiological investigations in psychiatry, it is obvious that investigators, depending on their theoretical perspectives, have played a 'numbers game' that either maximizes or minimizes the amount of mental illness that allegedly exists by changing the criteria used. If mental illness is viewed as the presence of a clearly established disabling condition, then obviously the estimate of its occurrence will be conservative. In contrast, if psychotic conditions, psychosomatic complaints, anxiety, and unhappiness are viewed as part of the same continuum, then it is evident that a large proportion of the population can be characterized as having some form of mental illness.

A study of Midtown, Manhattan (32), given great notoriety in recent years, estimated that approximately one-quarter of the more than 1,600 respondents between 20 and 59 years of age who were surveyed were impaired. Only 18·5 per cent of the total respondent group was evaluated as well. Similar observations have been made in other studies (19). In a study of the prevalence of mental disorders in Kalamazoo County, Michigan (20), in which various data comparable to those collected in the Midtown, Manhattan, study were obtained, the investigators demonstrated that the rates of mental illness in Midtown, Manhattan, were not very different from the community studied but were inflated by the very lenient criteria used to establish the presence of mental illness. They observe the following:

> Our interpretation is that the differences in reported rates of untreated illness arise *primarily* from lack of agreement, stated or implicit, in the criteria used to establish the cutting-point between the sick and the well. The criterion used in the Kalamazoo community study appears to identify only the extremely ill and to underestimate total prevalence. The Baltimore procedures seem to focus on a more broadly conceived spectrum to mental illness, though they, too, admit some underestimation. The rates reported for Midtown Manhattan are apparently the consequence of a very inclusive conception of mental illness (20).

Plunkett and Gordon (28), reviewing prevalence studies undertaken prior to 1960, note the widely varying percentages of the population found to be mentally ill from one study to another. The range in the proportion of populations found to be mentally ill varies from less than 2 per cent to as much as 33 per cent (pp. 62–8). Similarly, the Dohrenwends (8), in a review of the epidemiological literature, found estimates of morbidity varying from less than 1 per cent to as much as 64 per cent. The Dohrenwends illustrate clearly the unreliability of various absolute estimates of psychiatric illness. In comparing epidemiological studies carried out in 1950 or earlier with those done after 1950, they found widely varying rates of disturbance observed. The median rate in the studies after 1950 was more than seven times that reported for the earlier studies. Since it is inconceivable that population rates have changed very much, it is clear that these estimates are very unstable. Given the inadequacies of the measures of mental illness used in various surveys and field studies, it seems reasonable to use such measures primarily for assessing relative differences among population groups rather than treating the absolute levels of disorder reported as meaningful assessments (6).

As one considers more limited categories of mental illness, the possibilities of making some reasonable estimate of morbidity increase. When investigators use relatively limited concepts of schizophrenia, there is remarkable agreement in studies from one country to another that the prevalence of active cases varies approximately from one-quarter to one-half of 1 per cent of the population (33, 35). As the concept is expanded, of course, the prevalence rate is found to increase. From the point of view of public policy, questions concerning the prevalence of mental illness must be linked to decisions concerning the appropriate range of facilities that should be provided to the population. Once we have some conception of what conditions it is reasonable to treat, then we can estimate the extent of the problem in the population that we will have to deal with.

I think that I need not belabour the fact that the quality of case-finding and case-identification in psychiatric epidemiology is highly deficient. Treated hospitalization is too frequently confused with treated prevalence and sometimes with true prevalence. Even the more sophisticated studies rarely go beyond defining cases treated by clinics and psychiatrists, although there is general appreciation that a great deal of serious psychiatric pathology is seen in general medical practice (31), to say nothing about many conditions that are not seen at all. If one confines one's interest to a diagnosis such as schizophrenia, and has access to a relatively homogeneous population over some period of time, cases of treated prevalence may not be too unrepresentative of all cases, but this assumption is not necessarily correct (23). The psychiatric epidemiologist certainly faces serious practical problems in case-finding and he must often settle for less than is ideal, given the absence of definitive knowledge on the character of mental illness and the ambiguity surrounding many central issues in psychiatry. Yet it should be apparent that much of psychiatric epidemiology demonstrates no clear attempt to use even those techniques to estimate error feasible under present conditions of knowledge and research technology.

The influence of response sets on identified illness

The problem of selection of patients into treatment facilities has led to a variety of attempts to estimate the prevalence of psychological disorders in the general population. By estimating true prevalence it is possible to correct for the various cultural, social, and situational factors affecting distribution of care and pathways to treatment. In the absence of clear criteria establishing a reliable and valid definition of psychological disorder, these estimates, as noted earlier, tend to vary quite widely. Not only is it not clear as to what constitutes proper cutting points relevant to the psychiatric endeavour, but also the instruments used in such epidemiological investigations are subject to a wide variety of response biases that occur differently among varying segments of the population in question. These differential response tendencies, thus, may play havoc with comparisons of rates of disorder among various subgroups of the population (8).

Thus far, even in more sophisticated studies there has been a lamentable tendency to group together methods in a fashion that combines the disadvantages of each of them and yields an uncertain outcome. The advantage of a standardized interview by psychiatrists is that case-finding reflects clinical opinion in all of its complexity. The advantage of using standardized items in scales of illness or disability is that one can gain assurance that, having gathered the same information from all respondents, they will be scored in a standard and unbiased fashion. The use of psychiatrists to judge interview protocols involving standardized items is a peculiar mixture that may have a democratic flavour in bringing the disciplines together, but which frustrates each of the individual approaches (32). The data are not sufficiently rich for a typical clinical assessment and the use of clinical assessments on interview data frustrates whatever standardization becomes possible with developed indexes of psychological impairment.

One way of evaluating the validity of screening instruments is to test these instruments on known groups. If items can effectively discriminate between known 'ill groups' and known 'non-ill groups' they are said to have a certain validity. Since few items are fully discriminatory, our indexes are inevitably made up of imperfect items. By using a large number of items, each of which has some discriminating power, one can then limit the amount of error in the over-all assessment, since the judgement depends on the score on all items and not on any single one. The total scale is, therefore, more stable than its constituent items, and the more discriminating the constituent items, the better the scale. Using this approach to validating an instrument, it is then bizarre to include in one's scale items which have failed the test of ability to discriminate, because clinicians think these items are important nevertheless. There is no reason why such items cannot be used in a separate assessment on a clinical basis, but it is not sensible to combine these two types of items in a single scale so that they cannot be differentiated and scored separately.

The use of field interviews that involve screening questions for psychiatric disorder has some real advantages, but it also raises some grave problems. To the extent that one scores the specific responses provided by the respondent, one faces a large possibility of deception and distortion which may be intentional or a consequence of differential response tendencies (22). In recent years there has been considerable research on differential tendencies toward yea-saying and nay-saying, toward endorsing extreme as compared with intermediate alternatives and toward responding to the social desirability aspects of particular items. All of these possible biases become factors that demand consideration in the development of screening instruments and in their use in field studies.

In every situation of assessment—whether in a clinical context or in the field—the characteristics of the respondent may affect what is reported and what is withheld, the emphasis given to one set of experiences and symptoms or another, and the manner in which these are expressed and communicated. We know, for example, that many persons do not report accurately symptoms or conditions that are embarrassing to them or that involve social stigma (24). Thus, persons may withhold reporting not only psychiatric symptoms of particular kinds but also such symptoms as those involving venereal disease, female difficulties, etc. Since there are wide subcultural differences among people, there are varying thresholds among different respondents for describing particular kinds of symptoms.

One of the difficulties with using symptoms as indicators of morbidity is that they may have very different meanings in various subcultural groups. Aggression expressed by a person socialized in a group context that discourages and punishes aggression is qualitatively different from aggressive behaviour that is learned and is culturally normal. The values underlying the expression of aggression differ among social classes and ethnic groups, and in building instruments for assessment of mental disorder in populations it is wise to evaluate each of the indicators used from the perspective of the possible cultural biases they introduce.

Given the state of psychiatric work, as long as we use diffuse definitions of mental disorder we will be unable to distinguish social values and social risks from pathology. Drinking and smoking are both potentially harmful and both satisfy psychological needs, but only drinking is usually defined within a psychiatric perspective. Although psychiatrists tend to see the use of narcotics as within their orbit, as expansive as they have been in other respects, they have as yet not defined risky behaviour such as parachuting, car racing, motor-cycling, or skiing as appropriate areas for psychiatric concern. Heavy use of alcohol and experimentation with narcotics are forms of behaviour that provide certain satisfactions but also involve certain risks. But this is also true of parachuting and skiing, and it is not clear to me why some of these risky behaviours but not others are seen as pathological. It is not difficult to demonstrate that all of these activities involve risks that are above and beyond usual risks. Moreover, each of these activities is differentially distributed by class, ethnicity, and other social categories. The tendency to view the Irish lower-class drinking pattern and the Negro adolescents' experimentation with narcotics as potentially pathological but not the 'equivalent' attempts to find kicks and excitement characteristic of the more affluent classes suggests perhaps some biases inherent in psychiatric epidemiology.

The concept of impairment which is increasingly used in psychiatric epidemiology is really quite ambiguous. There are persons who suffer from profound and abundant psychophysiological indications but who function extraordinarily well in their social roles and obligations. To what extent are these people impaired in contrast to others who suffer no pain and lose no sleep but are dull, inefficient, and forgetful? A certain amount of anxiety is associated with high aspirations and it may enhance both motivation and excellence. There is a tendency, for whatever reasons, for many field studies to concentrate on much that is trivial and unimportant. If epidemiology continues to move away from the traditional nosological categories of psychiatric interest, it will have to focus on dependent variables that are more valid and less ambiguous than 'psychiatric disorder' or 'impairment'. For as long as we concentrate on such amorphous measures, we shall be doing no more than arbitrarily setting standards that have uncertain meaning. We are in a situation similar to those who study physical growth. The definition of who is tall or who is short is arbitrary and trivial. Physical measurement becomes meaningful when it alerts us to harmful conditions in the organism which can be remedied.

It might be useful if epidemiologists focused on the distribution of high-risk behaviours in the population in contrast to focusing on 'psychological disorder'. If risk was defined in terms of behaviours with an above-average chance of morbidity or mortality, we could define a universe of items that are not so closely tied to a particular system of social values. Indeed, we may end up studying a very heterogeneous universe of habits and behavioural patterns. Included among high-risk behaviours might be such diverse patterns as heavy smoking, driving after use of alcohol, dangerous sport activities, extreme depression, paranoid delusions, neglect of children, careless driving, suicide attempts, hunting, and carrying concealed weapons. Moreover, such behaviours as marijuana use, political protest, non-incapacitating psychophysiological symptoms, and low self-esteem might be excluded.

Obviously, epidemiological investigation depends on the issue in question and, indeed, different projects are motivated by very different issues. Given a clear criterion for hospitalization, if we wish to estimate the probable need for psychiatric beds, we do not measure psychiatric disturbance in general; nor do we use crude definitions of pathology if we seek to determine the causes of schizophrenia. In improving our case-finding definitions, we can move in very different directions and we should probably try a variety of approaches. There is much value in improving the traditional nosological categories that psychiatrists have used by refining and systematizing the criteria defining them, increasing their reliability, and obtaining a clearer picture of their validity when applied to non-hospitalized populations. It is evident that by careful standardization of interview procedures and by a clearer statement of criteria defining diagnostic categories, reliability can be substantially improved in many areas. We also must move in completely new directions in re-conceptualizing just what it is that we mean by 'pathology' and 'impairment', and how they can be distinguished from transient and situational reactions and arbitrary value judgements.

As Crandell and Dohrenwend (5) have suggested, there is also danger in using physiological indicators for assessing psychiatric distress. Many of the indicators used in screening devices are symptomatic of physical disease as well as mental disturbance, and if physical illness associated with these symptoms is not equally distributed through various social groupings, these may bias the estimated relative rates of psychiatric impairment in these groups. Moreover, some groups are more likely than others to report psychiatric disturbance in physiological terms, and to the extent that screening instruments depend on such indicators, possible biases are introduced in estimated relative rates.

Response tendencies are reflected not only in what people are or are not willing to report but also in the expressive mode with which they respond. Clinicians often assess the seriousness of symptoms by the concern they arouse and the distress which appears evident. But persons in varying groups express distress differently, and cultural values may prohibit or encourage the expression of pain (22, 38). It is, therefore, not uncommon for clinicians to confuse objective symptoms with the manner in which they are communicated (1, 37). Very anxious patients may appear calm and patients who communicate evident distress may have relatively minor symptoms. We obviously need to develop better approaches for separating response tendencies from pathology *per se*.

We also must face the possibility that response tendencies and pathology are like two sides of the same coin and that they cannot be effectively differentiated. In the experience of pain, for example, the subjective definition of the phenomenon is a powerful factor and inseparable from the physical sensation (2). Similarly, it may be that the manner in which certain psychological feeling states are perceived and defined affects their course and consequences, and thus the cultural pattern itself can be viewed as dysfunctional and pathological. From one perspective, delinquent behaviour and alcoholism are harmful whether they constitute learned subgroup patterns or personal deviant adaptations. But these two types of deviancy require very different modes of intervention and approach and any serious epidemiologist concerned with isolating causes and preventive measures would treat them separately. In short, even if one takes the position that some subcultural patterns are more conducive to health and adaptation than others, the approach to such variations is very different from the one taken in reference to patterns of behaviour which are the product of biological defects, disorganized and disrupted socialization, and those which violate group norms and expectations.

Although field studies of psychological disorder are becoming less rare, much of epidemiological effort will continue to be focused on cases that come into treatment. Considering the wide range of practitioners who deal with psychiatric disturbances, it is likely that greater emphasis will be given to case-finding by general practitioners, psychologists, and possibly even clergymen. In the emphasis on treated conditions, two important points must be kept in mind. First, every effort must be made to distinguish between the conditions which afflict persons and the forces that may differentially bring afflicted persons into treatment (21). Second, we must be aware that as we move from more serious incapacitating conditions to more common forms of psychological disorder, these selective forces bringing persons into treatment are better predictors of case-finding than pathology itself (25, 30). To summarize current evidence, it appears that for conditions such as schizophrenia and other more profound conditions, selective forces only exert a modest influence on case-finding in Western countries. When emphasis is placed on conditions of a neurotic type and personality disorders, it becomes apparent that social variables have profound effect on which cases come into treatment, and the cases that come into treatment are biased samples of all such cases in the population.

In recent years there has been a substantial amount of research on the various influences that motivate people to seek treatment or that stimulate others to see that some form of intervention is initiated (22). Two concepts frequently used to describe such forces are the concepts of illness behaviour and of societal reaction. When we speak of illness behaviour we refer to the various cultural, social, personal, and situational forces that lead to the varied ways in which persons perceive, evaluate, and act in reference to bodily indications. Illness behaviour thus encompasses such areas as pain recognition and expression, receptivity to the use of medical and psychiatric services, tendency to seek release from normal obligations and responsibilities, and the like. In contrast, the concept of the societal reaction refers to the differential responses others make to a person's behaviour, and this concept constitutes part of the social psychology of labelling. It is commonly recognized that behaviour is formed not only by an individual's motives and prior conditioning, but also by social expectations of others and by the manner in which social definitions help organize opportunities for social interaction. Such labelling processes can vastly expand a person's opportunities and potentialities, but they also can restrict and retard them.

The various factors affecting whether individuals themselves seek care or whether others in the community define them as requiring intervention of some kind are encompassed in the following ten categories, which are discussed in detail elsewhere (23):

(1) The visibility, recognizability, or perceptual salience of deviant signs and symptoms.

(2) The estimate made of the present and future danger likely to follow such signs and symptoms.

(3) The extent to which symptoms disrupt family, work, and other social activities.

(4) The frequency of the appearance of deviant signs and symptoms or their persistence.

(5) The tolerance threshold of those who are exposed to and evaluate the deviant signs and symptoms.

(6) Available information, knowledge, and cultural assumptions and understandings of the evaluator.

(7) Basic needs that lead to autistic psychological processes.

(8) Needs competing or interfering with illness responses.

(9) Competing possible interpretations that can be assigned to the symptoms once they are recognized.

(10) Availability of treatment, physical proximity, and psychological and monetary costs of taking action (including not

only physical distance and costs of time, money, and effort, but also such costs as stigma, social distance, feelings of humiliation, and the like).

The influences that affect factors such as visibility, tolerance, and the degree of annoyance and disruption caused by particular patterns of symptoms are not necessarily correlated with the degree of pathology as viewed from a medical or psychiatric perspective. Indeed, such factors may vary widely in different cultural contexts, among the different social strata, and under varying community conditions. Yet it is these factors characterizing illness behaviour and the societal reaction that may determine the public recognition of illness and the provision of psychiatric and social assistance.

Influence of differential theoretical models on the style of epidemiological investigation

Case-identification is crucial to any epidemiological investigation, but in psychiatric epidemiology the definition of a case is very much dependent on the theoretical assumptions made by the investigator. Most European epidemiologists have worked with the traditional nosology of psychiatric diagnosis, and they have made concerted efforts to sharpen and standardize the criteria used for identifying particular psychiatric syndromes (34). Implicit in this attempt is the assumption that each psychiatric diagnosis implies a different underlying syndrome which must be described separately in terms of its course, aetiology, and approaches to treatment and rehabilitation. Although these epidemiologists concede that they have not yet identified unitary psychiatric conditions, and that a diagnosis such as schizophrenia may include a variety of different underlying syndromes, their efforts are motivated by a concern with sharpening the diagnostic process and locating specific underlying conditions explaining the bizarre and dysfunctional responses indicative of 'mental illness'. Those applying the traditional medical perspective to psychiatric disorder also tend to believe more strongly than other psychiatrists in the importance of physical and genetic factors in explaining these syndromes.

In contrast, there has been a strong tendency among many American psychiatrists to assume that the same aetiological forces underlie a wide range of psychiatric disturbance, and that it is useful to regard psychiatric functioning as a continuum from wellness to illness (18). Implicit in this approach is the belief that if more common forms of psychological distress are dealt with effectively, chronic psychoses can be prevented (4). The implication of this position is that there is no fundamental difference between schizophrenia and a variety of other psychological disturbances, but rather that the differences that exist are the product of the manner in which social and environmental forces impinge upon persons. Given this point of view, it is both reasonable and convenient to treat psychological disorder or disability as a single variable. And it is well known that larger efforts in psychiatric epidemiology in the United States have concentrated on generalized disorder rather than on specific syndromes.

Fundamental to the difference in perspective between these two positions is the issue of whether more common forms of psychological distress are stable or transitory, and whether the same forces affecting the prevalence of such disorders similarly affect the occurrence of the diagnostic categories which European epidemiologists give emphasis to. This is an extremely difficult question which cannot be answered unequivocably at this time, but more recent investigations are providing us with very important data on these points.

Although there is evidence that persistent pathology may result from exposure to extreme situations of stress, these data do not allow one to determine whether such persistent responses are typical among persons who are not predisposed to mental illness (8). There are, however, many studies that suggest that the prevalence of psychotic conditions is not influenced greatly by changes in environmental conditions, although psychological symptoms in general may vary with social stress. Rates of hospitalized psychoses are relatively similar among developed Western nations and they do not appear to vary greatly from one period to another (33). Despite radically changing policies used by psychiatrists within the armed forces which resulted in substantial alterations in reported rates of admission for psychoneurotic conditions, there is impressive consistency in the rates of psychoses requiring hospital care (12). These rates do not appreciably vary from peacetime to wartime, nor were they very different in World War II as compared with World War I. And there is various evidence showing no significant change in rates of psychotic breakdown under extreme combat conditions or exposure to bombing attacks (29). There is no evidence that rates of psychotic breakdown significantly increase in civilian populations exposed to bombing attacks or to other major stresses (14). Also, contrary to popular conceptions, there is no adequate evidence that urban living, the hectic pace of modern life and the like, affect rates of psychotic breakdown (9, 26). An excellent and carefully executed study of the rates of mental illness over a century in Massachusetts found no support for the contention that the occurrence of the psychoses was increasing in modern life (10).

In contrast, there is evidence that intensive stress over time results in psychological disturbance that usually subsides when the stressor is removed (8, 16). Moreover, stress may precipitate breakdown in functioning among persons with satisfactory adjustment who suffer from psychological vulnerabilities. Obviously, conditions other than the psychoses may lead to severe problems incapacitating individuals and affecting their abilities to carry out their social roles. Current evidence suggests, however, that these stress responses and transient symptoms are qualitatively different from responses more characteristic of the psychoses.

In any case—whatever our interpretation of the evidence may be—we must remain aware of the fact that our assumptions concerning the resolution of this basic issue affect the kinds of epidemiological investigation we are likely to undertake. My own guess is that while it may be very useful in planning services to have some idea of the range and magnitude of psychological disturbance in the population in general, we will have to make much more specific studies of different syndromes or patterns of symptoms if we are to contribute to a sophisticated understanding of basic processes underlying mental illness.

Differential measures of outcome in epidemiological investigation

Epidemiological investigations are often concerned with determining outcomes of mental illness not only for the purpose of studying the effects of varying intervention alternatives, but also to determine the definition of continuous cases which constitute an important component of the prevalence rate. Historically, there is much evidence of administrative abuse of mental health statistics for the purpose of achieving certain political ends (7). If experience is an indicator of what we should strive for, then it is plain that the assessment of the effects of varying mental health policies must be based on more than admission and release statistics that are easily manipulated by administrators for their own purposes; it must be based on clinical and social indicators that have been standardized and that characterize the outcomes for patients and the community in very specific terms.

There are at least three ways in which the effects of varying mental health policies can be characterized and evaluated:

(1) We can measure the subjective response of patients to various kinds of policies, i.e., whether they feel they have been helped, whether they feel they have less symptoms and are more able to cope, and the like.

(2) We can attempt to measure their performance and the quality of life they live after exposure to varying public policies.

(3) We can attempt to assess the consequence of various policies in terms of economic and administrative costs, i.e., does one policy lead to an equivalent outcome at less cost than another.

Each of these forms of evaluation are important and compatible.

In recent years the most striking change in mental health policy resulted from the changing use of psychiatric hospitalization. Many patients who would have been retained in mental hospitals in the past are now returned to the community after a relatively short period of stay. Most patients leave the hospital within a year, and the average length of stay is no more than a few months. In short, if the quality of mental patient rehabilitation is to be measured by the ability to remain in the community or the ability to return to it, the consequences of new mental health policies have been profound.

The perceived importance of keeping the patient in the community rather than in the hospital is evident from the report of the Joint Commission on Mental Illness in the United States and the legislation that has followed the Commission's report. It is contended that:

> The objective of modern treatment of persons with major mental illness is to enable the patient to maintain himself in the community in a normal manner. To do so, it is necessary (1) to save the patient from the debilitating effects of institutionalism as much as possible, (2) if the patient requires hospitalization, to return him to home and community life as soon as possible, and (3) thereafter to maintain him in the community as long as possible. Therefore, aftercare and rehabilitation are essential parts of all service to mental patients, and the various methods of achieving rehabilitation should be integrated in all forms of services, among them day hospitals, night hospitals, aftercare clinics, public health nursing services, foster family care, convalescent nursing homes, rehabilitation centres, work services, and ex-patient groups. We recommend that demonstration programs for day and night hospitals and the more flexible use of mental hospital facilities, in the treatment of both the acute and chronic patient, be encouraged and augmented through institutional, program, and project grants (15).

Returning the patient to the community is, in itself, no panacea if the quality of the patient's life and functioning cannot be improved. Given the great emphasis placed in Western value systems on effective performance and carrying one's own weight, it is not surprising that some persons view residence in the community as a 'good' in itself, independent of the quality of the patient's life outside the hospital. But this poses the issue too simply. For if the patient is sufficiently aberrant and disturbed, his residence in the community may pose innumerable difficulties for his family and others, and result in a general outcome far inferior to good institutional care. It is therefore essential that we understand more thoroughly what happens to the mental patient outside the hospital—the extent to which difficulties occur and how they are handled. Intelligent planning of administrative decisions and community services must depend on a firm understanding of the true consequences of varying public policies. It is in this area that informative data are most difficult to obtain.

The studies by Brown *et al.* (3), by Grad and Sainsbury (11), and by Pasamanick (27) provide a warning to us that changing administrative policies and modifying psychiatric procedures to a limited extent are not enough to launch a successful community care programme (13). In our enthusiasm for change we would do well to heed this warning. Grad and Sainsbury have come to believe that the essential ingredient of relieving burden is to assure relatives that help is available and something could be done. This 'sense of control', which Grad and Sainsbury imply, has been recognized in other areas of investigation as an important factor mediating stress response in difficult circumstances (17). The investigators further suggest that for 'certain types of patients' in 'certain social circumstances' home care may leave the family with many more problems than hospital care. But since psychiatric patients are not a homogeneous group, it is necessary to learn in far greater detail what patients are more likely to flourish in home care and under what social and psychological circumstances. Moreover, it should be clear that if the public's responsibility is to be met, it is necessary not only to change administrative procedures, but to guarantee that adequate services are available to help to relieve the consequences of such changes and to promote effective functioning of patients outside the hospital.

Conclusion

In short, it is clear that much remains to be done in the development of psychiatric epidemiology. Particular attention must be devoted to the difficult problem of case-identification, and we must be cognizant of the underlying assumptions and points of view that focus our inquiries in one direction or another. We must also strive to develop more sensitive and sophisticated measures of outcome than has been typical in psychiatric epidemiology. Although it would be foolish, given the uncertainty of our knowledge, to insist that epidemiological efforts move in one direction rather than in another, it is prudent to be constantly aware of the extent to which our concerns are influenced by prevailing value systems that vary from one investigator to another.

REFERENCES

1. Bart, P. B. (1968). 'Social structure and vocabularies of discomfort: what happened to female hysteria?', *Journal of Health and Social Behavior*, **9**, 188–93.
2. Beecher, H. (1959). *Measurement of Subjective Responses* (New York: Oxford University Press).
3. Brown, G. *et al.* (1966). *Schizophrenia and Social Care* (London: Oxford University Press).
4. Caplan, G. (1964). *Principles of Preventive Psychiatry* (New York: Basic Books).
5. Crandell, D. L., and Dohrenwend, B. P. (1967). 'Some relations among psychiatric symptoms, organic illness and social class', *Amer. J. Psychiat.* **123**, 1527–38.
6. Davis, J. A. (1965). *Education for Positive Mental Health* (Chicago: Aldine).
7. Deutsch, A. (1949). *The Mentally Ill in America* (New York: Columbia University Press).
8. Dohrenwend, B. P., and Dohrenwend, B. S. (1969). *Social Status and Psychological Disorder: A Causal Inquiry* (New York: J. Wiley).
9. Fried, M. (1964). 'Effects of social change on mental health', *Amer. J. Orthopsychiat.* **34**, 3–28.
10. Goldhamer, H., and Marshall, A. W. (1953). *Psychosis and Civilization: Two Studies in the Frequency of Mental Disease* (New York: Free Press).
11. Grad, J., and Sainsbury, P. (1966). 'Evaluating the community psychiatric service in Chichester: results', *Milbank mem. Fd Quart.* **44**, 246–77.
12. Group for the Advancement of Psychiatry (1960). *Preventive Psychiatry in the Armed Forces with Some Implications for Civilian Use* (Report No. 47, Topeka, Kansas).
13. Hoenig, J., and Hamilton, M. W. (1967). 'The burden on the household in an extramural psychiatric service', in *New Aspects of the Mental Health Services*, 612–35, eds. Freeman H., and Farndale J. (New York: Pergamon Press).
14. Janis, I. L. (1951). *Air War and Emotional Stress* (New York: McGraw-Hill).
15. Joint Commission on Mental Illness and Health (1961). *Action for Mental Health* (New York: Science Editions).
16. Kern, R. P. (1966). 'A conceptual model of behavior under stress', *Technical Report 66–12* (George Washington University Human Resources Research Office).
17. Lazarus, R. S. (1966). *Psychological Stress and the Coping Process* (New York: McGraw-Hill).
18. Leighton, A. (1967). 'Is social environment a cause of psychiatric disorder?', in *Psychiatric Epidemiology and Mental Health Planning*, eds. Monroe, R. R. *et al.* Psychiatric Research Report 22 (April), 337–45.
19. Leighton, D. *et al.* (1963). *The Character of Danger* (New York: Basic Books).

20. Manis, J. G. *et al.* (1964). 'Estimating the prevalence of mental illness', *American Sociological Review*, **29**, 84–9.

21. Mechanic, D. (1962). 'Some factors in identifying and defining mental illness', *Ment. Hyg.* (*N.Y.*), **46**, 66–74.

22. —— (1966). 'Response factors in illness: the study of illness behavior', *Social, Psychiatry*, **1**, 11–20.

23. —— (1968). *Medical Sociology: A Selective View* (New York: Free Press).

24. —— and Newton, M. (1965). 'Some problems in the analysis of morbidity data', *J. chron. Dis.* **18**, 569–80.

25. —— and Volkart, E. H. (1960). 'Illness behaviour and medical diagnoses', *Journal of Health and Human Behavior*, **1**, 86–94.

26. Murphy, H. (1961). 'Social change and mental health', in *Causes of Mental Disorders: A Review of Epidemiological Knowledge, 1959*, 280–329 (New York: Milbank Memorial Fund).

27. Pasamanick, B. *et al.* (1967). *Schizophrenia in the Community* (New York: Appleton Century Croft).

28. Plunkett, R. J., and Gordon, J. E. (1960). *Epidemiology and Mental Illness* (New York: Basic Books).

29. Reid, D. D. (1961). 'Precipitating proximal factors in the occurrence of mental disorders: epidemiological evidence', in *Causes of Mental Disorders: A Review of Epidemiological Knowledge, 1959*, 197–216 (New York: Milbank Memorial Fund).

30. Scheff, T. (1966), 'Users and non-users of a student psychiatric clinic', *Journal of Health and Human Behavior*, **7**, 114–21.

31. Shepherd, M. *et al.* (1966). *Psychiatric Illness in General Practice* (London: Oxford University Press).

32. Srole, L. *et al.* (1962). *Mental Health in the Metropolis: The Midtown Manhattan Study* (New York: McGraw-Hill).

33. Wing, J. (1967). 'The modern management of schizophrenia', in *New Aspects of the Mental Health Services*, 3–28, eds. Freeman H., and Farndale, J. (New York: Pergamon Press).

34. —— *et al.* (1967). 'Reliability of a procedure for measuring and classifying "present psychiatric state" ', *Brit. J. Psychiat.* **113**, 499–515.

35. Wing, L. *et al.* (1967). 'The use of psychiatric services in three urban areas: an international study', *Social Psychiatry*, **2**, 158–67.

36. Yerushalmy, J. (1966). 'On inferring causality from observed associations', in *Controversy in Internal Medicine*, 659–68, eds. Ingelfinger *et al.* (Philadelphia: Saunders).

37. Zola, I. (1963). 'Problems of communication, diagnosis and patient care', *J. med. Educ.* **10**, 829–38.

38. —— (1964). 'Illness behavior of the working class', in *Blue-collar World: Studies of the American Worker*, 351–61, eds. Shostak, A., and Gomberg W. (Englewood Cliffs: Prentice-Hall).

Discussion

E. ESSEN-MÖLLER

University of Lund, Sweden

My discussion will be confined to a narrow part of a vast subject, namely, the collection and evaluation of psychiatric data in a population. Here I feel that the trend of opinion differs in some respects from my own, so I am interested in trying to bring out wherein the differences lie.

Of course, the same standards of recording and diagnostics should be used for the population as for clinics and hospitals, if one wants to compare the panorama of mental disorder prevailing in them. The diagnoses pertaining to the population should have approximately the same meaning and foundation as those of the clinic. Much has been written to explain why this is difficult to achieve. One factor is concerned with time: in order to find the disordered individuals, the investigator has to examine the whole population. I think this was the incentive for the division of labour which is now widely practised —the primary recording is entrusted to laymen equipped with a questionary, and so the time of the psychiatrist or clinical psychologist can be spared for subsequent screening of the records and diagnosis of the categories. My first point of discussion relates to this solution of the time problem.

I think it useful to remember in this connection that, in the ordinary clinical work with patients, recording of information does not necessarily come first and then diagnosis afterwards. This may be so with disciplines which depend essentially on laboratory findings but in psychiatry these processes are largely one and the same and take place simultaneously. Here the diagnostic attempts and the recordings both begin at the moment of the first encounter with the patient and continue incessantly throughout subsequent consultations and treatment. The process involves continuous attention to the informant's responses in the widest sense, to contents as well as to emotional reactions and subtle behaviour. What is seen and inferred, of course, depends on training and skill. Both sets of observations, conceptual and emotional, may carry diagnostic value. However, the observations also automatically influence the direction of the inquiry, which is thus subject to constant re-focusing, according to the nature of the responses. One might say that the diagnostic process, in its very formation, itself decides what will be probed for and looked for. This is what is meant by saying that recording and diagnosis are essentially one and the same and simultaneous.

There is, of course, nothing new or specific about all this, and in principle it is the same with any social intercourse or conversation, but its implication in our present problems may at times be underrated. For one thing, the natural process is interfered with when recording and diagnosis are carried out by different persons. Secondly, the use of lay recorders and questionaries has effected, more or less inadvertently, a standardization and thus a reduction of information. The diagnoser will be in doubt whether there are more cases than those indicated by the records and whether, except for the most conspicuous disorders, the positive indications are correct. To be sure, the questionaries may be highly perfected and may include ratings of behaviour, but the many directions a free interview might have taken can hardly be foreseen in a questionary, and not much variation in behaviour can be observed in a person whose answers are limited more or less to yes or no. In these respects, a free examination should be superior.

In other words, the preferable method of approaching clinical standards so far appears to consist in having resort to a clinically experienced person—who ought to be familiar also with the dialect and attitudes of the population—so that he himself can do all the recording, screening, and diagnosing with a free rein. I believe this is the best method and perhaps somebody will be able to prove it.

Yet there continue to be serious doubts about this method. Some of these doubts run as follows: We know how different is the training within different schools of psychiatry, and how different accordingly are the diagnoses and even the records. Even with identical training, individual psychiatrists or clinical psychologists differ slightly in their outlook. How then can we feel sure that the findings of different authors are comparable? It has been argued that the less one knows about theory the better one is able to observe and record events objectively. If this were so, since no clinician—not even an eclectic—will be without some prejudices from theory, we would be back again to the use of lay recorders and questionaries.

You will have noticed that the above reasoning in defence of a standardization is no longer concerned with the technical problem of saving the clinician's time. It is rather a matter here of a standardization for the purpose of conceptual agreement, and this is the second topic I want to discuss. The problem is not specific to epidemiology; it applies to clinic and population alike. In fact there is a recent trend for using standard questionaries and records even in clinical psychiatry. One might speculate on the further development of this: perhaps ultimately the clinician will be exchanged for a computer:

This indeed might be an effective if somewhat unexpected way of bringing clinical and population diagnoses to an equal level!

Please do not think I am sceptical of all kinds of standardization whatever. It need not be underlined how great is the importance of a wider agreement on the use of terms and, still more perhaps, of our knowing how terms are used in different quarters. The ongoing work of promoting such understanding and standardization, internationally as well as within national groups, is no doubt worthy of support. Important also is that kind of standardization or structuralization—again more technical in nature—which serves to remind the interviewer of the areas to be covered but does not interfere with his method. I wish we had used a little bit more of this sort in our own population study.

However, I continue to be afraid, rightly or wrongly, that claims for a standardization of concepts may go too far. Doubtless each separate school is 'able to observe and record events objectively'. The trouble is they will focus on different events. Accordingly, while comparisons between the findings of different schools is desirable, it would probably be better to compare them on the level of their final conclusions, as it were, rather than on that of records and diagnoses. For example, each separate school may have studied the incidence of mental disorder in the different social groups of a population. While a crosswise comparison of detailed frequencies within corresponding groups would be of little use because of the different concepts and terminology, much more information might be derived from comparing the trend throughout the groups of the first population with the corresponding trend within the other. But again, in order for such a comparison of second order to reflect realities, it is mandatory, I imagine, for the different devices of approach to be carried out in an optimal way, that is, freely. Confinement to shared or standardized concepts would certainly yield much of interest but would probably also unduly limit the possibilities of more penetrating research.

It is similar with another kind of limitation and here I arrive at my third and last point. It has been recommended by our introductory speakers that we should restrict the scope of our interests and observations in order to prevent psychiatric epidemiology from becoming vague and non-committal, and that we should avoid dealing with trivialities such as, for instance, tallness and shortness of stature. In a way I can see the reason for this and I agree that many practical questions would perhaps be most readily solved by focusing on some particular point and nothing else. But as soon as clinical or theoretical problems are involved, I believe there may be danger in such restriction. Throughout psychiatry there are endless transitions and dilutions which cannot well be dispensed with in clinical research, and much of such research needs to be done directly in the population. An example will show what I mean.

In respect of stature as well as intelligence, we know that variation has two overlapping components: one is the so-called natural variation, while the other is caused by abnormal genes or exogenous lesions which have interfered with an otherwise normal development. I would not say that this is trivial, and still less so if the principle is also valid for variations in personality. If it is valid, one might be able to distinguish, in some part at least, the apparently completely normal individuals—those who represent the 'natural' dimensions—from those whose personality, even if socially healthy, displays some trait however mild which might indicate a hypothetical pathologic interference.

This is what we attempted to do in our population. Every inhabitant was seen and briefly interviewed by a psychiatrist and in addition outside information was obtained, and thus a crude classification of everybody was made, even of the socially healthy. We followed here the diagnostic principles of our teacher, the late Professor Sjöbring, who worked at them for a lifetime. Ten years later, Hagnell was able to demonstrate in the same population that the interim first incidence of mental disorder, mostly mild, was significantly higher in some of our tentative 'pathologic' groups and, correspondingly, substantially lower in the supposedly 'natural' variants. Even if this outcome might be interpreted somewhat differently, it constitutes, together with the underlying line of thought, one more reason why I have come to believe in a broad and free range of research.

To conclude, I believe that a clinical approach to psychiatric epidemiology is preferable to certain standardized procedures, invented for the purposes of time-saving, uniformation, or concentration. This may appear a conservative position, but what has to be conserved, after all, is freedom!

REFERENCES

ESSEN-MÖLLER, E. *et al.* (1956). 'Individual traits and morbidity in a Swedish rural population', *Acta psychiat. neurol. scand.*, Suppl. 100.

—— (1961). 'Methods and techniques for obtaining comparability in field studies of mental disorders', in *Field Studies in the Mental Disorders*, discussion, ed. Zubin, J. (New York: Grune and Stratton).

HAGNELL, O. (1966). *A Prospective Study of the Incidence of Mental Disorder* (Lund: Svenska Bokförlaget).

SJÖBRING, H. (1963). *La Personnalité: Structure et Développement* (Paris: Doin) (English edition in preparation).

2

FOLLOW-UP STUDIES INVESTIGATING CHILDHOOD DISORDERS

Follow-up studies investigating childhood disorders[1]

LEE N. ROBINS
Department of Psychiatry
Washington University Medical School
St. Louis, Missouri, U.S.A.

The State Commissioner of Mental Health asks himself: 'If 50,000 babies are born this year, how many child psychiatrists, school social workers and psychologists are we likely to need to care for them in the next two decades?'

The mother of a new-born baby asks herself: 'Is there anything I must do or avoid doing if I want to prevent the appearance of behaviour problems? My husband drinks too much, will my baby be affected? If my baby is colicky and fussy at three months does this mean he is going to be a problem child? If we live in a bad neighbourhood will my son become a delinquent?'

The parent or doctor of a child who already has behaviour problems wants to know answers to a different set of questions: 'How long are these problems likely to continue? Are there any treatments likely to shorten their natural course? Do they presage some serious illness in adult life?'

Follow-up studies of psychiatric disorders in children have addressed themselves to the following issues: the risk of the appearance of pathological behaviour in populations of children; the different rates of risk in children with different kinds of heredity, social settings, early behaviour, and maternal practices; the duration of the disorder; the evaluation of treatment; and the predictive value of problem behaviour in childhood for adult adjustment. This paper will discuss what special advantages the follow-up study brings to collecting information about these issues, and will illustrate the contribution made by some of the more interesting follow-up studies in the literature.

---

[1] This work was supported in part by USPHS Special Fellowship MH-36,598 and National Institute of Mental Health Research Grant MH-07126.

Characteristics of follow-up studies

The concept essential to the follow-up is that it involves measures at two or more points in time. The *kinds* of things measured at these two points may be the same or different. We may, for instance, study how many children aggressive at time 1 are still aggressive at time 2, i.e., the *same* set of behaviours measured twice. But we may also study how many children with sleep disturbance at school entry are failing school at age 12, two different kinds of behaviours. The *people* studied at time 1 may or may not be the same as the people at time 2, although there must be a stated relationship between them. If we investigate whether children from broken or impoverished *homes* at time 1 will have behaviour disorders at time 2, we are in fact studying connections between the behaviour of their *parents* and their own behaviour.

Follow-up studies can resemble laboratory experiments if during the elapsed time some planned intervention has taken place (for instance, treatment). But they may also treat elapsed time as a period of spontaneous change and development only—and thus become explorations of natural histories.

In this paper, two restrictions will be added to the broadest use of the term follow-up. First, the time interval must be of sufficient length so that there has been an opportunity for an important amount of change to occur. Second, the measurements at the two points in time must be independent of each other. These strictures rule out studies comparing the diagnoses of children on entering and leaving a treatment facility, since changes in diagnosis over such a short interval probably more often reflect changes in the *amount of knowledge* about the child acquired by his physician than changes in the state of the child himself. They also rule out studies in which a teacher's evaluation at the beginning of the school term is found to correlate with the child's grades at the end of the term, since the fact that the same person (the teacher) has provided both the initial rating and the final grade makes the two measures fail the test for independence.

The special assets of follow-up studies

One reason a follow-up is preferable to a cross-sectional study, when we wish to know which children develop psychiatric disorder, its susceptibility to treatment, and its long-term prospects, is that the follow-up technique avoids retrospective falsification and selective recall. If, in a cross-sectional study, the mother of a disturbed child is asked about family variables present before the child showed symptoms, her account is likely to be distorted in at least some of the following ways:

(1) It will revise the past so as to 'make sense' of the relationship between past events and the child's behaviour. For instance, the mother may report that the grandmother's death occurred before the child began having nightmares, when in fact the nightmares had started months earlier.

(2) She will remember things that mothers of normal children would not, because she is seeking explanations for her child's problem. For instance, she may recall that the child was frightened by a dog, or that a distant relative had a somewhat similar kind of problem. While these events *did* occur, they may occur just as frequently in the history of normal children, but are too trivial to be recalled by a mother who is not concerned over her child's behaviour.

(3) Guilty about her child's problems, she may exaggerate her own inadequacies as a parent, or to avoid personal blame may paint a blacker picture of the father's irresponsibility than she would if she did not have a problem child.

(4) Like the mother of the normal child, she may simply misremember important events or confuse the history of the problem child with that of another of her children. In follow-up studies, the information about the past was *collected* in the past. It is, therefore, more likely to be accurate, because it deals with recent events and because it cannot be altered to explain or justify what happened later.

One advantage that has been claimed for the follow-up study is that it can discover the *causes* of behaviour disorders. In fact, the follow-up study contributes to the understanding of cause in only a limited way. *No* non-experimental study, whether cross-sectional or longitudinal, can *demonstrate* cause. The only way to do so is by experimentally manipulating a variable that you think produces behaviour disorders and then looking to see whether you have in fact changed the rate. But non-experimental studies can help to decide which factors are *probable* causes. To argue that some particular circumstance is a probable cause of psychiatric disorder, you must show that when it occurs to children free of disorder, their risk of developing a disorder is increased. The difficulty is in showing that the circumstance actually was present *prior* to the onset of the disorder. For instance, while parents are often hostile and unloving toward delinquent children, this does not necessarily mean that the parents' attitude influenced the children's behaviour. We need to know whether the parent was hostile *before* the child showed any kind of abnormal behaviour, or whether the hostility is simply the parents' response to the child's provocative acts. A *retrospective* or *prospective* study that starts with delinquent children, who may have been discipline problems at home long before they were known to the courts, obviously has trouble in deciding whether the parent was already rejecting and disapproving before the child *ever* did anything to provoke the parent. But can follow-up studies of non-delinquents do better? If a study starts with problem-free children, some of whom have hostile parents, and follows them to see which ones develop problems, and if the children with hostile parents *do* develop behaviour problems at a disproportionate rate, it appears probable that the hostility of the parent has played a part. But it is difficult to find such children to study. One must identify the children at that critical period when they have *already* been exposed to the factor that we think causes difficulties, but *before* they have reacted to that exposure by showing problem behaviour. Obviously, then, we can use follow-up studies to test causal hypotheses in quite limited circumstances. While follow-up studies solve *some* of the dilemmas of research into the causes of behaviour problems, they are by no means a panacea.

A final area in which follow-up studies are particularly valuable is in estimating what proportion of the population of children ever develop behaviour disorders. Suppose we were to use a cross-sectional study of 7th graders to make such an estimate. We would miss all the children whose disorders were severe enough so that they never entered 7th grade—children in hospitals, correctional institutions, special schools, and ungraded rooms, as well as children whose reckless and impulsive behaviour led to fatal accidents. On the other hand, if we were to select for follow-up at age 12 a representative sample of very young children chosen before psychiatric disorder can influence their accessibility, we will need to account for all our cases, and our figures for the rate of development of psychiatric disorder before puberty will include children dead, out of school, or retarded in school, along with those who progress normally.

The main advantages of follow-up studies, in short, are:

(1) That information about early life is more accurate.

(2) The temporal order between environmental factors and the development of psychiatric disorder can sometimes be ascertained.

(3) A more complete sampling of the population can be achieved.

These advantages, however, do not accrue automatically. Many of the follow-up studies we will discuss have not taken full advantage of the potentials of the follow-up design. Nonetheless, some important results have been achieved, which could not have been achieved with cross-sectional studies.

A sampling of follow-up studies of children

A. RATES OF PSYCHIATRIC DISTURBANCE IN CHILDREN

The proportion of children who will ever develop psychiatric disorders can be estimated by follow-up in two ways. One method is to choose children at birth, follow them until the age designated as the end of childhood, and see how many developed problems at any time in the interval. A second way is to select well children of various ages (or children born at different periods), follow them for a fixed period of time—a period long enough so that *some* of them will have reached the age designated as the end of childhood, and see how many new cases of behaviour disorders emerged at each age-level. With proper statistical treatment the risk of the development of a disorder within the whole age span of childhood can then be calculated.

The Berkeley Growth Study is an example of the first method. Every third child born in Berkeley, California, over an eighteen-month period was followed at yearly intervals until he reached age 14 (33). The children were evaluated at each age with respect to the presence or absence of 39 symptoms. While this study does not attempt to label children as having or not having psychiatric disorders, the proportion of children having each symptom has been reported by sex and age, and the average number of symptoms present at various ages computed. We learn, for instance, that less than 5 per cent of children have more than 10 of the 39 symptoms at age 11. We also learn what symptoms tend to occur together and what symptoms are rare at particular ages. For instance, poor appetite, a rare complaint at age 14, occurs so frequently in little girls of 6 (40 per cent) that it probably has no psychiatric significance at that age. Running away from home, on the other hand, occurs in fewer than 10 per cent of children at every age and so may be an indicator of serious problems whenever it occurs. Similar findings *could* have been obtained with a cross-sectional rather than a longitudinal design by sampling children of all ages and comparing the proportions at each age showing each symptom. But then we would have had to *infer* that the reason fewer 11-year-old boys steal than do 6-year-olds is because stealing drops out of the repertoire of behaviours with aging. In a follow-up study, using the *same* boys at different ages, we *know* that this happens. We do not have to wonder whether our 11-year-olds just happen to be less antisocial children than our 6-year-olds.

Two Swedish studies have collected data appropriate for the second kind of study. Unfortunately, both present only the figures for the number of new cases found without estimating how many new cases could have been expected if all children had reached the upper age limit.

In one of these studies, Jonsson and Kalvesten (27) selected, over a 9-year period, one-sixth of all Stockholm's birth records of boys born on the 15th of the month (0·5 per cent of all boys born during this period) and then located these boys at ages 7 to 16. The rate of problem behaviour was calculated from information obtained from school records; interviews with teachers, mothers, and fathers; examination by a psychiatrist; and psychological tests. One-fourth of the boys were found to have problems serious enough to warrant referral for treatment, and 2 per cent had been institutionalized. Excess activity was found in 6 per cent; 5–10 per cent were defiant and aggressive; 20–5 per cent oversensitive; and 18 per cent had reading or writing difficulties. Serious delinquent behaviour (truancy, vagrancy, stealing) was reported for 2–3 per cent.

In the second Swedish study, Hagnell (22) followed an entire community ten years after an initial psychiatric survey. While this study concerned adults as well as children, results for those under 10 at initial study provided information about children. Of children under 10 who were free of psychiatric disease in 1947, 7 per cent of both boys and girls developed psychiatric illnesses by 1957. This incidence for children was less than half the incidence of new cases of psychiatric illness among adults well in 1947 (17·6 per cent). The most common diagnosis for children was what the author called 'childhood neurosis', which apparently corresponds to 'acting out' or antisocial behaviour in the U.S. Boys (6 per cent) developed this pattern more frequently than girls (1 per cent). Girls' problems were more commonly what we call neurotic: fatigue, depression, anxiety, and mixed neurosis. None of the children developed a psychosis. During the ten-year period, only 2 per cent of children had consulted a psychiatrist and only 0·5 per cent had been in a psychiatric hospital. The proportion of those becoming ill who were hospitalized was lower for children than for adults (7 per cent versus 12·5 per cent). Apparently, then, children less often develop psychiatric illness than do adults, and their illnesses may be less severe, since when sick they less often require hospitalization.

B. CHILD POPULATIONS WITH HIGH PSYCHIATRIC RISK

Predictors of psychiatric disorder in children can be sought in characteristics of the social environment or in early characteristics of the child himself. Indexes of disorder found associated with early environmental or personal characteristics include delinquency, coming to psychiatric attention, and poor school performance. Mental subnormality is also strikingly related to social factors but this will not be discussed here.

1. *Socio-economic factors and parental performance*

The most ambitious study ever undertaken concerning the effect of socio-economic and family factors on the life chances of children is the follow-up study of a sample of 5,362 children born in a single week of 1946 in England and Wales (12, 13, 42). These children have been repeatedly restudied from birth to age 20. Findings published thus far report family and social status factors affecting the child's risk of delinquency, of passing examinations for entrance into and progress through secondary schools, and of neurotic symptoms.

Low social status, as indicated by the father's having a manual rather than a white-collar job, was found to be related to official delinquency and to the appearance of aggressive behaviour in school, to low scores on school achievement tests, to *declining* scores on school achievement tests when scores at ages 8 and 11 are compared, to critical comments by teachers such as 'poor worker' or 'lazy', and to failure to enter schools for the intellectually able, even when achievement scores warranted admission. That is, lower-class children have more antisocial behaviour, *achieve* less, and do not even reap the rewards of what they do achieve. No class differences were found in the number of neurotic symptoms reported (bed-wetting, nightmares, nail-biting, thumb-sucking, abdominal pain or vomiting).

This study sheds light on *why* lower-class status is associated with behaviour problems. When the working-class families of delinquents were compared with other working-class families, the parents of delinquents were found to have less education; they had made less use of health services available; they had more often quarrelled and separated. In short, they appeared less-adequate parents. If less-adequate parents are more common in the lower class than in the middle class, this could explain why the lower class produces a higher proportion of children with behaviour disorders. It is reasonable that inadequate parents *should* be more common in the lower classes, because inadequacy would show itself in an inability to hold a job or to obtain jobs other than those at the lowest level, as well as in poor performance as a parent.

Even if the higher proportion of inadequate parents in the lower classes accounts for much of the excess of disturbed children among the poor, low social status itself might be an additional factor. To show status is *not* important in its own right, it must be shown that (1) an improvement in social status does *not* lead to a decrease in children's behaviour problems unless there is comparable improvement in family adequacy, and (2) that inadequate parents are as likely to produce children with behaviour problems in middle-class homes as in lower-class homes.

The first question—does improvement in standard of living lead to improved behaviour—was answered negatively by a study of Negro slum children whose families were moved to a spacious, clean, new housing project (66). The children who moved into the project showed no greater gains in school achievement than did a matched group of children who stayed in the slum, although there was a slight decline in school absences among the project children. There is a question, of course, as to whether improved housing is enough to constitute a rise in social status. It is hard to imagine another *kind* of improvement in social status, however, which can occur without the parents' showing competence in work performance.

The second question, does inadequacy in the family produce equally severe behaviour problems in children in middle-class families as in lower-class families, was answered affirmatively by a study of the behaviour of three generations (49). It was found that having antisocial parents or grandparents (parents or grandparents who as adults had been arrested, drank excessively, failed to work regularly, deserted or neglected the children, beat the children or spouse, or had extra-marital sexual relations) considerably increased the chances of a son's dropping out of high school and appearing in police or juvenile court records. This relationship between problem behaviour in a boy and his relatives appeared in *both* white-collar and blue-collar families. Indeed, when both had antisocial relatives, boys in white-collar families had no advantage over boys in blue-collar homes with respect to graduation or arrests.

2. *Psychiatric illness in parents*

Persons treated for psychiatric illness may be especially likely to have children with disorders, either because they are particularly inadequate parents or because the child inherits their psychiatric disorder. Unfortunately, most studies reporting on psychiatric illness in the offspring of psychiatrically ill parents have been conducted after the offspring are already adult. It is not possible to tell from these reports whether or not the disorders reported for the offspring began *in childhood* or later. Some problems found *must* have been visible in childhood, especially mental deficiency and antisocial behaviour, since we know that these rarely, if ever, first appear in adulthood.

Most studies investigating disorders in the offspring of psychiatrically ill parents have been more interested in the effect of a genetic than a child-rearing factor in psychiatric illness. By contrasting the offspring of parents with differing diagnoses they show that psychiatric disorders tend to run 'true to type' within families. One study sought hospital records for the offspring of parents both of whom had been hospitalized with either schizophrenia or a depressive illness (31). When both parents were schizophrenic, 23 per cent of their offspring had been hospitalized for schizophrenia and none had been hospitalized for depressive illness. When both of the parents had had an affective (depressive) illness, only 8 per cent of their offspring had been hospitalized for schizophrenia and 19 per cent had had a depressive illness. When one parent was schizophrenic and the other depressed, the rates of the two diseases in their children were almost equal—10 per cent schizophrenic and 13 per cent depressed.

But disorders other than those of their parents also appear in the children of the psychiatrically ill. Are such disorders more common in the children of the psychiatrically ill than in the general population? If so, does this indicate that genetic inheritance is not specific with respect to diagnosis? Or does the excess of disorders unlike the parents' show the environmental effects of living with a psychotic parent or of suffering separation from the parent when he or she is hospitalized?

Two studies have shown that disorders unlike the parents' do occur more frequently in the children of schizophrenics than in the children of parents not known to be psychiatrically ill. One study compared the children of patients diagnosed *dementia praecox* with the children of patients seen in a medical clinic (6, 7). Since most of these children were still young at follow-up, the number developing behaviour disorders is not a final figure. However, a comparison of the *kinds* of disorders shows not only more cases of schizophrenia among the children of schizophrenic parents (6 per cent of the children over 16 with a schizophrenic parent versus less than 1 per cent of the control children over 16 years) but more cases of conduct disorder as well (11 per cent versus 2 per cent for school-aged children). A study of adult children separated from their mothers in early life and placed in foster homes or institutions (25), similarly showed not only an excess of schizophrenia in the children of schizophrenic mothers (11 per cent versus none in the control children) but also an excess of antisocial behaviour (15 per cent convicted of a felony versus 4 per cent of the control children) and mental deficiency (8·5 per cent versus none of the control children). This second study indicates that the excess of disorders other than schizophrenia in the children of schizophrenic parents cannot be attributed to living with the psychotic parent nor to the experience of separation from the parent, since in this study *all* children had been early separated from the mother, whether or not she was psychotic. While this study might then seem to suggest that a *genetic* component may be reflected in diverse kinds of pathology, this inference cannot safely be made because no information is available about the father of the child. If schizophrenic women more often have children by antisocial or mentally defective men than do other women whose children are removed from them, the high rate of non-schizophrenic pathology in the children may well reflect the genetic contribution of the father.

3. *Children's early behaviour as a predictor of later problems*

Most studies following the behaviour patterns of children have been concerned only with normal development, not with the appearance of psychiatric disorder. This was a wise choice, since the small samples used made the chances slim of turning up many psychiatrically ill children. A few studies, however, do relate early behaviour to the development of problems.

The Berkeley Growth Study, discussed earlier as contributing to our knowledge of the frequency of various kinds of symptoms found in children of various ages, has also exploited its repeated assessment of symptoms in the same children to learn at what age high symptom levels tend to persist and to learn which symptoms are transient and which are lasting (33).

Most children's problems seem age-specific. That is, a problem present at one age will not be found later. The symptoms which were found to be exceptions to this rule were destructiveness, demanding attention, sombreness, jealousy, shyness, and excess reserve. When these were present at age 6 or 7, they were also likely to be present at 13 or 14. But while most other specific symptoms change from one age to another, children with *many symptoms* at one age tend to have many symptoms later as well. This consistency in the over-all level of disturbance does not appear, however, until age 6 or 7. Before then, symptoms have little predictive value for later adjustment.

The fact that levels of symptoms are not good prognosticators before age 6 seems consistent with the findings of a study of 136 normal children first observed shortly after birth and followed by observation and parent interviews regularly thereafter. As infants, they were evaluated along nine behaviour scales. Before the age of 7, 29 of these children were referred and accepted for psychiatric help (56). By age 9, 42 were given psychiatric care (63). This unusually high rate of psychiatric referrals apparently occurred for two reasons: first, the children came from New York Jewish professional families, a group which every study of the utilization of psychiatric facilities finds unusually interested in and accepting of psychiatric intervention; second, the psychiatrist doing the research freely offered services as a recompense for the family's co-operation. Although it is unlikely that the children seen had behaviour disorders of the severity seen ordinarily in psychiatric practice, they had caused enough concern to their parents that a consultation was sought. And the psychiatrist agreed that the problems were not simply normal behaviour for their ages. When an attempt was made to correlate psychiatric treatment with the nine behavioural attributes scored in infancy, little could be found in the observations of early behaviour of patients to distinguish them from the less-troubled children. In the first year of life, only one of the nine attributes (high activity level) was significantly related to pathology; none of the variables scored in the second year; and only one score (high intensity) in the third year.

A follow-up of children in nursery school (65) indicates, however, that behaviour by age 3 or 4 may have predictive value. Immaturity, getting along poorly with other children, and eccentric behaviour noted in the nursery school teacher's records were related to receiving mental health services before completing high school. Failure to get along with peers in nursery school was related to trouble with peers and teachers later on and to childhood neurotic symptoms. Eccentric behaviour in nursery school was also related to neurotic symptoms later.

Three studies of school-aged children confirm the Berkeley Study's observation that behaviour after age 6 is indeed a very good predictor of later adjustment problems. These studies sought predictors of delinquency and school achievement in teacher's comments and evaluations by other students. All three studies (8, 23, 42) agree in showing that delinquents-to-be do poor school work, challenge the teacher's authority, and are unpopular with their schoolmates. In the British study (Mulligan), pre-delinquents were 'under-achievers' by age 8. Later they disobeyed teachers, quarrelled and fought, cheated, and played truant. In Denver (Conger), by 4th grade pre-delinquents were resentful of authority, getting along badly with classmates, day-dreaming, and doing poor work. In 'River City' (Havighurst), the children chosen as most aggressive by teachers and classmates in 6th and 7th grades later became delinquent, particularly if they also had school failures. Two of these studies agree as well that neurotic symptoms ('nervousness' in Denver; stammering, nail-biting, vomiting, bed-wetting, thumb-sucking in England) are unrelated to later delinquency. (Comparable data are not presented in the River City study.) Such neurotic symptoms neither predict delinquency nor predict that a child will not be delinquent. Delinquency and nervousness are unrelated but not incompatible.

School achievement, unlike delinquency, was predicted by *both* neurotic symptoms and aggressive behaviour. In England the achievement test scores of children with neurotic symptoms were found to deteriorate between the ages of 8 and 11. Since achievement scores are important determinants of which children the teachers will nominate for admission to schools for the academically talented, 'grammar schools', having symptoms in early childhood does predict that a child will not be admitted to the 'grammar schools' and consequently will end his schooling early. In River City, more than half (57 per cent) of the children among the most aggressive in 6th and 7th grades failed to finish high school, compared with only 17 per cent of the least aggressive.

4. *Traumatic experience and later problems*

A large literature has been devoted to the discussion of whether or not maternal deprivation is followed by behaviour disorders in childhood (67). The belief that maternal deprivation was important was fostered by the observation that children in orphanages often seemed to be suffering from a lack of spontaneity, low intelligence, and rather constricted ability to make relationships with others. The interpretation that such abnormalities in institutionalized children result from experiencing separation from the mother is no longer so widely accepted. Such behaviour may be a consequence of inadequate care in the institution. Or it may have existed prior to separation from the mother, in some cases caused by the mother's neglect or maltreatment prior to the child's removal. Separation from the mother is so severely disapproved in our society that it rarely occurs in the absence of serious pathology in either the mother or the child. When Bowlby (4) studied children re-entering school after prolonged hospitalization for tuberculosis, a cause for separation unrelated to psychiatric disorder in either the child or his mother, little or no difference could be demonstrated between the children who had been in a sanatorium and control children in their classrooms.

One kind of traumatic experience—abuse at the hands of the parents—has been found to be associated not only with very serious psychiatric problems for children, but also with serious medical problems as well (15). Among 50 abused children, defined by their being found at hospital admission to have multiple bone injuries at various stages of healing, who were followed a year and five months to ten years later, 8 were found to have died and 5 to have been permanently institutionalized in hospitals for the mentally defective. Among the 20 available for interview, half had IQs below 80, half had speech problems, and 40 per cent had 'emotional disturbance'. Children who had been removed from their homes showed better physical and intellectual development than those remaining, suggesting that when maternal behaviour is sufficiently abnormal, being deprived of the mother is an advantage. It is possible that some of the children had mental defect or brain damage prior to being abused, defects which may have made them likely to incite abuse through their unresponsive or demanding behaviour. But since removal from the mother improved outcome, it is reasonable to suppose that at least part of the disastrous findings at follow-up resulted from the parental abuse, along with the neglect and poor nutrition that often accompanied it.

C. CAN PSYCHIATRIC DISORDER BE PREVENTED?

Beginning with Cyril Burt's study of the delinquent in 1925 (5) and continuing with the studies by Healy and Bronner (24) and the Gluecks' history-making *Unraveling Juvenile Delinquency* in 1950 (20), it became obvious that the families and social environment of delinquents were unfortunate. Boys who became delinquent had had few 'good influences' in their lives. It was an easy step from this observation to experiments in delinquency prevention by offering good influences and attempting to counterbalance the influence of 'bad companions' and of parents who were uninterested or rejecting.

The results of these efforts have been consistent if disheartening. No experiment in delinquency-prevention has succeeded. The techniques used have been those of counselling of children and their parents, psychotherapy, encouragement to participate in wholesome recreation, offering of friendship, good examples, and understanding. The treatment has been applied at various ages: to boys around age 11 (36, 44); to adolescent girls (37); to children from kindergarten to 6th grade (62); and to boys just entering elementary school (9). Since each of these studies provided a carefully matched control group to allow comparisons between children for whom prevention was attempted and children for whom it was not, there is every reason to have confidence in the negative results.

While none was able to prevent delinquency, all the studies demonstrated that the *predictors* of delinquency were correct. Those children expected to become delinquent did, in fact, do so more often than children for whom delinquency was not anticipated, although prediction was far from perfect. The predictions were much better than chance whether the predictors used were family type, the parents' assessment of the child's behaviour, home visitors' judgement of the family, or teacher's assessments and school records.

Efforts to prevent childhood disorders other than delinquency are common, but have seldom been carefully evaluated. Prenatal clinics hold parents' meetings, to help prepare the parents psychologically for handling the prospective baby. Many parents' and teachers' associations show and discuss 'mental health' films, or invite psychiatrists to answer questions. But there have been few attempts to learn whether these efforts do in fact reduce the incidence of psychiatric disorder in children. One study (17) contrasted, over a three-year period, classrooms in which the mothers have been offered group therapy sessions and mental health films followed by discussion programs, with classrooms in the same school not exposed to the program. These efforts could not be demonstrated to have reduced the incidence of psychiatric symptoms in the children or to have improved teacher ratings of children's behaviour.

D. CAN BEHAVIOUR PROBLEMS BE SUCCESSFULLY TREATED?

If current techniques for preventing psychiatric disorder in childhood are inadequate, treatment might still be effective.

To decide whether or not a treatment has helped, one thinks first of assessing the child's status before and after treatment to see if there has been improvement. But it is not simple to evaluate the results. Psychiatric disturbance often has a fluctuating course. Children usually come to professional attention because of the sudden onset of a new symptom or the worsening of existing symptoms. If seeking treatment results from a temporary exacerbation of symptoms, a few weeks or months later the level of symptoms will usually have receded—whether or not treatment was obtained. Treatment is valuable only to the extent that *more* children have improved or have improved *more permanently* than could have been expected if they had not been treated. To decide whether treatment has 'worked', then, we need to know how much improvement over the same time interval would have occurred in untreated children with similar complaints, and we need to know how many of both treated and untreated children continue improved some years later.

There has been considerable argument about how to obtain the necessary base-line figures for how much improvement can be expected without treatment. One method used is to assess the improvement of children assigned to the waiting-lists of clinics. A careful review of the studies that compared treated and untreated cases referred to child-guidance clinics (32) finds little support for the idea that treatment makes much difference. Improvement rates of two-thirds to three-quarters are regularly reported for both treated and untreated populations studied immediately after treatment, and rather high relapse rates are reported in treated children followed a number of years later.

Questions have been raised about whether these negative findings do not result from differences in the initial degree of illness in the treated and untreated groups. For instance, emergency cases may be accepted for treatment and less severe cases assigned to waiting-lists. Or perhaps the children assigned to the waiting-lists were not really untreated—they may have sought help elsewhere.

Perhaps an even more basic problem than whether waiting-lists constitute proper control groups is whether it is reasonable to evaluate treatment by studying improvement rates in a total clinic population in the first place. No one would think of evaluating the effectiveness of medical treatment by seeing how many attenders of a general medical clinic are well or improved a year or two later, while comparing them with people who approached the clinic for treatment but were not seen. We know that some patients of a medical clinic will have a self-limited illness, such as influenza, from which *all* will have recovered. For them the relevant question is not whether they are improved a year later, but whether the course of the disease was milder as a result of treatment. Other patients, those with arthritis for instance, have presently incurable illnesses with spontaneously occurring remissions and exacerbations. The relevant question for such patients at follow-up is neither cure nor improvement, as measured by current symptom level, but rather whether treatment has increased the proportion of the follow-up interval during which the patient has been able to function effectively.

For other syndromes, the goal for therapy is to prevent further progress of the illness.

Yet many of the studies reviewed by Lewis have failed to specify the disorders being treated or what the goals of treatment were, as well as failing to ascertain whether the control group was really comparable. But these failings do not seem to account for the poor showing made by treatment. Studies dealing with specified problems and using clearly defined goals (preventing recidivism among delinquents) have not been able to demonstrate greater effectiveness for therapy.

While the total child psychiatric population shows a *high* rate of improvement, whether or not treated, one small subgroup of that population, psychotic children, shows just the opposite. Children diagnosed as psychotic have grossly abnormal behaviour, usually including the absence of relationships with their peers, an absence of speech or abnormal speech characterized by repeating questions verbatim and reversing pronouns (referring to themselves as 'you'), preoccupation with their own activities and failure to respond to interruptions, dislike for and resistance to changes in routine, stereotyped repetitive movements, failure to achieve toilet training, and temper tantrums. Such children *rarely* improve, whether or not they receive treatment. Among fifteen psychotic children given very intensive therapy, first in the hospital and subsequently as out-patients, only two children showed a little improvement, and even these two children were still grossly abnormal five years after initial diagnosis (14). These children had had almost all current therapeutic manoeuvres attempted—play therapy, milieu therapy, special education, speech therapy, drugs, as well as therapy with the parents, but all to no avail.

The California Youth Authority forestry camps have a modern treatment programme for delinquency, including counselling, work experience, and planned recreation. Early reports showed less recidivism for delinquents assigned to these camps than to traditional reformatories or to a prison. However, when boys whom judges deemed appropriate cases for the forestry camps were randomly assigned to the camps or reformatories or prison, the advantage of the forestry camps disappeared (40). Thus study of delinquents points out the great importance of making certain that experimental and control groups are truly comparable. The previously reported success of the forestry camps apparently had resulted from judges' selecting the good-risk delinquents for them.

Some studies have claimed success for modern facilities for delinquents based on the way attitude test questions are answered at the end of the period of incarceration. Such indirect methods of assessing 'delinquency-proneness' are not very satisfactory. One wants to prevent delinquent acts, not change questionnaire answers. It is still an unanswered question as to whether there is any correlation between such answers and later behaviour.

The effect of treatment in a child-guidance clinic on *non*-delinquent and *non*-psychotic children was studied by comparing the proportion improved over a two-year period with the proportion improved among untreated schoolchildren carefully matched with the clinic sample with respect to the nature and severity of their symptoms (60). This match for presenting symptoms was achieved by first screening the school population for children showing symptoms similar to those in the clinic, as ascertained by questionnaires administered to the parents of both groups, and then personally interviewing the parents of both groups until a good match for severity was located. At follow-up two years later, improvement rates for treated and untreated children were indistinguishable. As in less rigorously controlled studies, approximately two-thirds of both clinic and school problem groups had improved.

E. THE DURATION OF CHILDHOOD DISORDERS

We noted in the last section that childhood disorders seen in clinics, whether treated or untreated, usually improved with time. We also noted that delinquents and psychotic children, both treated and untreated, had much lower rates of improvement than the usual run of clinic children. Some studies have been interested in these spontaneous rates of improvement and in how much they differ between types of problem behaviour or disorders of differing degrees of severity. Since treatment was found not to be an important modifier of the course of childhood disorders, it will be no surprise that such natural history studies also find that most childhood disorders improve or disappear with time, but that the psychotic disorders and extreme antisocial behaviour tend to persist.

When teachers were asked to identify children who were 'problems' or emotionally disturbed, they selected 7·6 per cent of their classes, and twice as high a proportion of boys as of girls (10). When teachers were again asked to so designate children two years later, half of the group originally designated were renominated. Boys were somewhat more likely to be renominated than girls. We cannot optimistically report, however, that 50 per cent of all disturbed children recover within two years, since no attempt was made to locate children who had not remained in the school system or who had not been promoted both years. Almost certainly those not promoted and transferred to special schools had higher rates of problems than those progressing normally.

A similar study of 2nd to 5th grades (18) reported only 30 per cent of those originally identified as still emotionally disturbed four years later. This figure may be spuriously low because of the methods used to specify who was emotionally disturbed. Instead of being asked to nominate any children they considered problems, as in the preceding study, teachers were asked to pick the five most poorly adjusted children in their classes, and fellow-pupils were asked to nominate the three children they liked least. These nominations were combined with scores on a personality test to decide who was maladjusted. In classes with fewer than five children with serious problems, teachers had to nominate children with trivial difficulties, which might be expected to disappear. Even when children with serious problems were selected, their failure to be renominated later need not indicate a disappearance of problems. Perhaps enough children not previously nominated became seriously maladjusted during the interval to displace the child originally selected from among the teacher's 'worst' five or their classmates' most disliked three.

Follow-up studies of young offenders generally attempt to learn how many again commit a crime after an arrest. The Gluecks (19) followed 1,000 juvenile delinquents in three five-year periods. During the initial five-year period, at the end of which their average age was 19, all but 20 per cent had been rearrested. Similarly, among juvenile drug offenders in

California (46), all but 18 per cent had another arrest in the next four or five years. When the drug arrest was not the first juvenile offence, only 9 per cent failed to be rearrested. These studies emphasize the fact that serious antisocial behaviour, as reflected in a juvenile police record, is much more persistent than is the general run of childhood disorders noted by teachers or by child-guidance clinics.

When child-guidance populations are divided into diagnostic groups, it becomes clear that the one-third who do *not* quickly improve are predominantly the antisocial and psychotic children, rather than the ones with neurotic diagnoses. In a Detroit children's clinic (11), cases were divided into those with personality disorders (antisocial behaviour), psychosomatic disorders, and neurotic disorders (fears, eating and sleeping problems, etc.). The children were followed a year or two after referral, through telephone interviews with their mothers. The mothers reported most improvement for children originally classified as neurotic.

Similarly, in a follow-up study from the Worcester Clinic (61), neurotic symptoms were found to be associated with good outcome, while restlessness and inattention, symptoms commonly described in antisocial children, were associated with poor outcome.

A five-year follow-up of 72 adolescents in an out-patient clinic again found a good prognosis for neurotics (91 per cent unimpaired at follow-up) and a poor prognosis for antisocial adolescents who were diagnosed 'sociopathy' (35). All the sociopaths were moderately or severely impaired at follow-up. Similarly all the schizophrenics were still impaired five years later.

The same pattern is found for adolescents disturbed enough to require hospitalization. Two to five years after admission, the schizophrenics have the poorest outcome (19 per cent recovered and 23 per cent improved), the antisocial behaviour disorders (sociopathy) are next (38 per cent recovered and 22 per cent improved), while the neurotics do well (40–55 per cent well and 24–50 per cent improved) (1). Among patients diagnosed antisocial behaviour disorder, prolonged disturbance was predicted by the presence of *multiple* antisocial symptoms (stealing, violence, and truancy). With only an isolated antisocial symptom, no matter what its nature, adolescents had good prognoses.

Warren (64) classified his adolescent psychiatric in-patients as 'neurotic', 'conduct disorders', 'mixed neurotic and conduct disorders', and 'psychotic'. His psychotics presumably include patients Annesley and Masterson would have called schizophrenics and his conduct disorders are presumably similar to Annesley's and Masterson's behaviour disorders and sociopathy. If so, the findings of all three studies agree well. Among children younger than 15 at admission, Warren found at follow-up six or more years later that most (74 per cent) of the neurotics were well, while only a third of the conduct disorders had recovered. Almost all of the psychotic children were still sick at follow-up.

F. HOW SHOULD CHILDHOOD DISORDERS BE CLASSIFIED?

Lacking laboratory tests for psychiatric illness, the psychiatrist relies heavily on the past history of the illness in order to make a diagnosis. When the patient is a child, there is very little past history available. One does not yet know whether his illness has a remitting or deteriorating course nor what the full spectrum of symptoms will be when the disorder has been present for some time. Such an absence of history makes classification difficult. One obvious classificatory distinction one wishes to make is between disorders unique to children and those which are early forms of well-recognized adult disorder. Following children with disorders over a period of years makes it possible to learn *which* adult disorders do and which do not occur in children.

One adult psychiatric disorder about which this question has been asked is hysteria. Hysteria in adults is a disorder characterized by a great profusion of symptoms, typically including menstrual symptoms, anxiety symptoms, frigidity, pain on no known physical basis, and multiple surgical operations. It occurs primarily, if not exclusively, in women, and almost always begins before the age of 25. To learn whether it can occur in children, a follow-up study was done of all the children in a children's hospital who had been diagnosed hysteria and hypochondriasis, as well as those diagnosed mixed neurosis if their symptoms included any of the classical 'hysterical' symptoms: pain, vomiting, paralysis, amnesia, fits, urinary retention, blindness, trances, or aphonia (47). Of the 23 children examined a number of years later, only 4 were found to be diagnosable as hysterics. The 4 for whom the original diagnosis was valid were all girls who had presented more than 5 symptoms. They averaged 8·25 symptoms. Children later found *not* to have hysteria averaged only 3·53 symptoms. The diagnosis for children who did *not* have hysteria as adults was most frequently anxiety neurosis (8 cases). Four children were found to have no psychiatric disease. This study indicates that hysteria *can* begin before puberty but that most children who show a combination of nervousness and unexplained physical complaints will as adults still be nervous, but their physical complaints will either have been discovered to have some physical basis or will have dissipated without ever having been diagnosed.

Another diagnosis about which a question has been raised as to whether it occurs in children is schizophrenia. While some psychiatrists make a diagnosis of childhood schizophrenia in withdrawn children who show little or inappropriate emotion, others feel that this syndrome is quite distinguishable from true schizophrenia, and is, in fact, associated with brain damage.

Support for the occurrence of schizophrenia in childhood was claimed by a study in which a series of 120 pre-pubertal children diagnosed 'childhood schizophrenia' at Bellevue Hospital between the ages of 4 and 13, were located again when they were aged 11–20 (2). Fifty-seven per cent had received a diagnosis of 'schizophrenia' in another hospital after leaving Bellevue, and an even higher proportion were diagnosed schizophrenic at follow-up by the Bellevue staff themselves. While the author argues that the high rate of diagnosis of schizophrenia both by the Bellevue staff and by other hospital staffs shows that true schizophrenia does occur in children, this is not necessarily an appropriate inference. The Bellevue staff was interested in confirming its original diagnosis. And the other hospital staffs presumably had the Bellevue diagnosis in hand while making *their* diagnosis. No objective criteria for the diagnosis of schizophrenia were applied at either intake or follow-up. At intake, the only requirement appeared to be that the symptomatology invade every area of the child's functioning. The specific symptoms could be not only varied but at opposite ends of a continuum—from anxiety and withdrawal to 'over-adapted and relating to others too well'. It is difficult to decide the meaning of a follow-up study in which the original diagnostic criteria are unspecified and where no provision has been made to avoid contamination between the original diagnosis and the follow-up diagnoses.

More recent studies (14, 57, 58, 59) throw considerable doubt on the idea that children who are psychotic before puberty are really schizophrenic. Among 63 children first seen between the ages of 2 and 11 years and then followed ten years later by Rutter, none was found at follow-up to have the cardinal symptoms of schizophrenia, i.e., delusions, hallucinations, or paranoid ideas. This same failure to develop delusions or definite hallucinations had been reported in an earlier study of autistic children (28). Eaton and Menolascino found children originally diagnosed childhood schizophrenia indistinguishable clinically at follow-up five years later from children with known neurological defects at intake, and much more severely damaged than would be expected in ordinary 'simple schizophrenia' of adult life. In addition, the families of Rutter's psychotic children did not contain an excess number of schizophrenics, as families of adult schizophrenics do. The childhood psychotics were very predominantly male, while there is no excess of males among adult schizophrenics. On three grounds, then—symptomatology, family history, and sex distribution—the childhood psychotics did not resemble schizophrenics when they grew older. An alternative hypothesis to the theory that psychosis in early childhood is schizophrenia was suggested by the frequency with which these children developed fits and other neurological signs during the follow-up interval. The delayed appearance of neurological signs is presumptive evidence for brain damage, even in some children who show no neurological abnormality early in their illness.

Another contribution that follow-up studies can make is to distinguish between *primary* and *secondary* symptoms. Many child psychiatrists, noting that an absence of speech and total withdrawal from social interaction were characteristic of psychotic children, had inferred that psychotic (autistic) children *could* speak but did not do so because they had no interest in communicating with others. However, in some of the non-speaking autistic children Rutter followed the extreme social withdrawal *disappeared* as they got older but speech did *not* appear. This finding suggested to him that the absence of speech is *not* simply an expression of social withdrawal. Indeed, he wondered whether the social withdrawal is not rather a consequence of the child's inability to understand the communication of others and to communicate with them.

While disagreeing with Bender about whether psychotic children are schizophrenic, the Rutter, Eaton and Menolascino, and Kanner and Eisenberg studies all agree with her that the prognosis is very poor. In all studies, about half the children were institutionalized at follow-up. An evaluation of those out of institutions shows a poor level of adjustment with almost none employed. These studies, then, add to the studies of psychotic adolescents mentioned in the previous section, to show that when psychosis occurs in very young children, the prognosis is also very poor.

G. DO CHILDHOOD DISORDERS PORTEND ADULT PROBLEMS?

One of the common folk feelings about disliked childhood behaviour is that children 'grow out of it' before they reach adulthood. Research has shown that the folk feeling is correct in many cases but that there are important exceptions.

Research investigating the likelihood that the disorders of childhood will be carried over into adulthood have been of two types:

(1) Investigations which start with disturbed *children* and then locate information about their adult adjustment.

(2) Investigations which start with persons known to have psychiatric problems as *adults* and then search childhood records to learn whether they also had problem behaviour as children.

The second type of study may not at first appear to belong under the heading of follow-up studies, since the index reference point is the later period of time, just as it is in cross-sectional studies that ask the patient about his early history in an attempt to elucidate the causes of his present difficulties. But it differs from such retrospective studies in that the data about childhood were recorded during childhood and are thus not contaminated by knowledge that the subject would have adult psychiatric problems. Such studies do indeed, therefore, meet the criteria for follow-up studies delineated earlier: measures were taken at two different points in time, the time interval was long enough for significant change to have occurred and the measures were independent of each other.

Studies which begin with children as the point of reference are only extensions in time of the studies of the duration of childhood symptoms which we described in an earlier section. The children are merely followed beyond adolescence into adulthood. In this section, we will discuss only studies in which at least some subjects were 25 years of age or older at follow-up, so that stable adult patterns can be discerned. Our criteria for having had childhood disorder are—being known to psychiatrists, being known to the police or courts, and having evidence for learning or disciplinary problems in elementary school.

1. *Children known to psychiatrists*

Two studies have investigated adult outcomes of children hospitalized in psychiatric facilities. Masterson (34) followed 153 adolescents hospitalized in the Payne Whitney Hospital 5–19 years after discharge. He found outcomes for adolescents very similar to outcomes reported for hospitalized adults. That is, adolescents diagnosed schizophrenia had a better prognosis the older they were at onset, when there were symptoms of depressed mood and/or confusion, when the personality prior to the illness was good, and when there was marked improvement while in the hospital. (In the opinion of some researchers these results, in both adult and adolescent studies, might best be interpreted to mean 'when the patients were *mis*diagnosed schizophrenia'.) Psychoneurotics almost all had good outcomes. About half the sociopaths (antisocial personality) had good outcomes and those who improved had *not* had difficulties in school and had improved markedly while in the hospital. (Might these also be examples of misdiagnosis?)

Morris (41) also followed hospitalized children, but his group was younger (aged 4–15) when initially hospitalized and was confined to non-psychotic, non-brain-damaged, antisocial children of normal intelligence. This group is probably most comparable to the group in Masterson's study who were diagnosed sociopathic. Out of a group of 90, 47 were followed at least to age 26. Although not psychotic as children, one-quarter of these antisocial children were diagnosed schizophrenia as adults and one-third (including the schizophrenics) had been chronically hospitalized. Only one-quarter had a good adult social adjustment. Those who were well adults had shown least symptomatology prior to hospitalization. This study suggests that while *childhood psychosis* may not be an early form of schizophrenia, antisocial behaviour serious enough to lead to hospitalization does sometimes forebode adult schizophrenia.

Studies of the adult outcome of child-guidance clinic populations have shown the same striking patterns found in the more seriously ill hospitalized populations. Both in Dallas (38, 39) and in St. Louis (48), poor outcomes were largely confined to *antisocial* children. Antisocial children had not only more psychiatric hospitalization as adults, but also more difficulties with the law, with their jobs, with their families, and with social relationships of all kinds than either neurotic or control children.

The Dallas study found a later hospital diagnosis of schizophrenia most common in the 'ambiverts'—boys judged to be neither introverts nor extroverts. In the St. Louis study, schizophrenia was somewhat more common as the number of childhood neurotic symptoms increased (4 per cent of children with 2 or fewer, 6 per cent of those with 3–5, and 9 per cent of children with 6 or more), while the presence of antisocial symptoms did not change the risk. In a sense these findings agree, since patients in the St. Louis study with a high degree of *both* antisocial and neurotic symptoms (more than 6 of each) produced a higher proportion of schizophrenics than did any other group (11 per cent). Both studies agree that the shy, withdrawn personality, often thought to be predictive of schizophrenia, predicts neither schizophrenia nor other psychiatric illness, and that some antisocial children are schizophrenic adults.

The St. Louis study found that having a *father* whose behaviour was antisocial increased the risk that antisocial behaviour would persist into adulthood. Social class, broken homes, gang membership, on the other hand, had little predictive value. One important finding was that *no* childhood variables, neither the child's behaviour problem nor his family type, predicted adult neurosis. Indeed, many of the symptoms for which children are commonly referred to psychiatrists—tics, speech difficulties, shyness, fears, over-sensitiveness, nervousness, irritability, tantrums, and insomnia—occurred as often in children well as adults as they did in those who were later neurotic. These findings suggest that while adult antisocial behaviour has its roots in childhood behaviour disorders, adult neurosis for the most part does not. (An exception was hysterical neurosis, which could be predicted by both family patterns and behaviour.)

The later military history of boys seen in child-guidance clinics of several cities (52) was contrasted with the military history of a control group of randomly selected schoolboys. Child-guidance clinic patients were found to have had more rejections from service for behaviour problems and criminal records and fewer deferments because of dependency and essential jobs during World War II. Among boys inducted, the clinic patients received more dishonourable discharges and more honourable discharges for psychiatric illness or 'unsuitability', indicating that they had adjusted less well to life in service. Despite their higher rate of problems, most clinic patients (74 per cent) had uneventful military careers. It was found possible to predict from the clinic records *which* patients would fail in service. One of the best prognostic clues was poor relationships with other children, as reported by teachers, supplemented by evaluations by family members and clinic personnel (54).

2. *Children known to legal authorities*

Although there is a large literature on the prediction of recidivism among juvenile delinquents, a literature which has grown out of the needs of judges and probation officers to decide on the proper disposition of children in their custody, there have been few studies which followed juvenile delinquents long enough to judge either their long-range risk of arrests as adults or their adjustment in other areas.

The Gluecks followed 1,000 boys referred to the Judge Baker Clinic by the juvenile court until they reached an average age of

29 (19). For 848 information was available for each of three 5-year follow-up intervals. Only 13 per cent were found to have been entirely free of delinquency throughout the 15 years since their initial referral and 58 per cent had been arrested during the last 5 years (between the ages of 24 and 29, approximately).

While this study does not provide a control group, the number arrested seems extremely high. That it *is* very high is validated by studies in St. Louis. Two-thirds of the male white child-guidance clinic cases and normal controls had been known either to the police or juvenile court before the age of 17 (48). More than half (58 per cent) of those subjects known to the police as juveniles had a significant adult police record (sentenced for a non-traffic offence, or, if not sentenced, arrested at least three times for non-traffic offences). When the subjects had had *no* police contacts as juveniles only 14 per cent ever had a significant adult police record. A similar strong relationship between delinquency and adult arrest records has been found in a follow-up study of Negro schoolboys in St. Louis (Robins, unpublished). Two-thirds of the Negro men with a juvenile police record had a significant adult police record, as compared with only one-quarter of those free of juvenile offences.

Juvenile arrest history was a more powerful predictor of adult arrest records for both whites and Negroes than was any measure of family stability or social class. However, for whites, but not for Negroes, family pathology and lower class status in childhood did *add* to the risk of later arrests for delinquents. More than 4 out of 5 (84 per cent) of the white delinquents from lower-class problem families had significant adult arrest records.

A comparison of the military careers of juvenile delinquents with the careers of a random sample of schoolboys (55) found that about a quarter of juvenile delinquents were rejected by the services on 'moral' grounds, as compared with only 7 per cent of a random sample of boys. When delinquents were inducted, more than half had unsatisfactory service records, compared with only 13 per cent of the control group of schoolboys. This study shows that the bad adult prognosis for juvenile offenders is by no means limited to their criminal histories.

The widespread predictive power of juvenile delinquency appears also in a study of delinquent and non-delinquent children known to a child-welfare board in Sweden (43). Children were referred to the board because of bad homes, delinquent behaviour, or both. Children referred for *delinquency* had worse adult outcome than children from bad homes, not only in terms of their police records, but also in terms of marital disruptions, poor occupational success, and high rate of alcoholism. Like the St. Louis study, Otterström's study emphasizes that the presence of childhood behaviour problems is a more powerful predictor of adult outcome than is family pathology. Both studies also show that antisocial behaviour in childhood predicts a wide variety of adult difficulties.

3. *Children identified in school as showing problem behaviour*

In the St. Louis study, the very high level of adult success in the slum-dwelling control group, chosen for an absence of elementary school retardation and serious truancy, was in marked contrast to the outcome of the child-guidance clinic patients (48). Most of these control children, although from impoverished families, as adults had little crime, stable family lives, steady employment, rising social status, and little psychiatric disorder. Finding that an absence of elementary school problems predicted such good adult adjustment suggested that the *presence* of retardation and truancy in elementary school should predict adult *mal*adjustment. This prediction is being tested in a sample of Negro men. While analysis of data is not complete, the combination of being held back at least one semester in elementary school plus being absent more than 20 per cent of the time in five or more school quarters has already been shown to predict death before the age of 33 (11 per cent versus 2 per cent for the remainder of the sample) (50), problem drinking among those surviving to age 33 (32 per cent versus 20 per cent of the remainder of the population) (51), a serious adult criminal record (57 per cent versus 34 per cent for the remainder of the sample), and diagnosable adult psychiatric disorder (66 per cent versus 39 per cent for the remainder of the sample).

This study replicates in a non-patient, non-white sample, the follow-up study of white child-guidance patients which found school performance in children of normal intelligence a powerful predictor of adult adjustment and adult psychiatric status.

4. *The childhood histories of sick adults*

When childhood predictors of *rare* adult psychiatric disorders are the topic of interest, it is wasteful of research time and funds to start with a random sample of children. One would follow them into adulthood only to discover that too few have developed the disorder to allow statistical comparisons between their childhoods and the childhood of subjects without the disorder. One solution is to select children thought to have a high rate of *predisposition* to the disorder (as the St. Louis clinic population had to adult antisocial behaviour), but when the disorder is as rare as schizophrenia, for instance, a population of deviant children still will not produce enough cases. If one chooses a still more highly predisposed sample, as for instance the offspring of schizophrenic mothers, the schizophrenics one discovers at follow-up will not be a *representative* sample of schizophrenics. Most schizophrenics, after all, do not have a mother so diagnosed.

A solution to this problem is to select a representative sample of *adult* schizophrenics and then to locate appropriate records dating from their childhoods. A control group can be chosen from the same childhood records—for instance, the next name in the list of classmates or clinic attenders—or from the adult records, depending on the question one seeks to answer.

This technique of following cases 'backwards' has been used in a number of studies of patients identified in mental hospital records. In one study (26), the names of state hospital patients admitted from one metropolitan area were checked against the records of a child-guidance clinic operated during their childhoods by the Board of Education. Patients *not* located in the clinic records were assumed to have been free of significant maladjustment in childhood. Very impressive differences in the frequency of appearing in the childhood records were found among different diagnostic groups. Almost all of the patients with an adult diagnosis of schizophrenia (*dementia praecox*), sociopathy, or mental deficiency had been known to the clinic as children, but none of those with depressions and few of those with situational reactions or acute organic psychoses.

A somewhat similar study in England, starting with adult patients and locating psychiatric records made for them as children (45), again found that adults with personality disorders (sociopathy) more often had been seen as children than had those with depressions. (Rates of schizophrenia were not discussed, probably because this diagnosis was much less common in the acute psychiatric hospital studied than in the state hospital population.) This study also attempted to relate childhood symptoms to the adult diagnoses. Among patients seen as children, depression and anxiety in adulthood had been prefigured by neurotic symptoms in childhood (predominantly anxiety, depression, and somatic complaints). Adults with personality disorders (sociopathy) came predominantly from those children referred for antisocial behaviour (stealing, truanting, lying, fighting). This part of the study confirmed the St. Louis study's findings that antisocial behaviour in childhood was strongly related to the diagnosis of sociopathy in adults.

Male patients diagnosed schizophrenic in one hospital were searched for in the records of a child-guidance clinic in the same city (16). When a name was located, the next male patient of the same age was selected as a control subject. The purpose of this study was to see whether the symptoms of child-guidance clinic patients destined to be schizophrenic differed from the symptoms of other clinic patients. The pre-schizophrenics were found to have lower IQ scores, to have a greater number of symptoms and to have had a gradual onset of symptoms. They had more school failures and came from more disturbed families. Particularly striking symptoms were their inability to make normal social relationships, their frequent temper tantrums, and their description as listless. They did *not* show symptoms of psychosis in childhood.

Two studies sought school records of persons hospitalized with a diagnosis of schizophrenia and compared their records with records of matched controls in the same school. In one study (3) school personnel were interviewed about patients and control subjects using 'double blind' methods (neither the interviewer nor the school personnel knew which boy was hospitalized and which was well). School reports for patients differed from reports for controls by showing lower IQ scores, poorer grades, more deteriorating grades, and more dropping out of school. The patients also more often were described as careless, apathetic, depressed, having few social contacts, and being less athletic. Although poor adjustment in high school had been common among patients, about one-fifth had been extremely well-adjusted students.

The finding that pre-schizophrenics perform poorly in high school is open to at least two interpretations:

(1) That the disease affects school performance and social interaction long before frank delusions and hallucinations appear.

(2) That the disease occurs most frequently in people from culturally deprived backgrounds, backgrounds which prepare them poorly for school success as well as making them vulnerable to schizophrenia.

One study permits ruling out the second interpretation. It compares IQ test scores in the elementary school records of adult schizophrenics with the scores of their own siblings (29). Since siblings share both class status and family environment, differences in IQ could not be attributed to class or family differences. When compared with their *own* brothers and sisters, the future schizophrenics were found to have lower IQ scores. Correlations between their IQ scores and their siblings' were much lower than the correlation between siblings in the general population. These findings suggest a disease process already started in elementary school which depresses IQ and wipes out the correlations attributable to genetic factors and family environment found in normal siblings. Since similar findings did *not* occur between siblings when one had been diagnosed neurotic in a child-guidance clinic, this finding does not appear to apply to *all* psychiatrically ill persons. Schizophrenics were also found to have lower birth-weights recorded on their birth certificates than had their siblings (30). From this, we may infer that the pre-schizophrenic child is different at birth.

Conclusions

This paper has discussed some of the studies in English published prior to 1968 which meet the following criteria: the studies have dealt with children's disorders, broadly defined as reported symptoms, difficulties at school or with the law, or referral to a psychiatrist before the age of 18; measures have been taken at two or more points in time, at least one measure having been taken before the age of 18; the measures have been independent of each other; and more than a year has elapsed between the initial measure and the follow-up measure. An effort has been made to select examples of studies meeting these criteria from each major area in which follow-up studies have made a contribution. In areas in which several studies have been published, quality of the research has been the guiding factor in choosing which studies to present. Judgement as to quality was based on the size and representativeness of the sample, the clarity with which methods were reported, and the use of adequate control samples.

Although some of the less rigorous studies have thus been eliminated, the scientific qualities of the studies presented vary widely. It is perhaps a tribute to the contribution that follow-up studies can make that concordant findings appear despite major differences in the orientations of the researchers to one school of psychological thought or another and in the scientific rigour with which the studies have been pursued. We would argue that this concordance stems at least in part from the fact that follow-up studies require independent measures at two points in time, a requirement which reduces the possible bias of results in the direction of the researcher's preconceptions.

There are many questions about childhood disorders which could be answered by follow-up studies but to which satisfactory answers have not yet been found. Among these questions for which we have sought answers in existing studies, the least well covered are the incidence of disorders among children, the duration of childhood disorders, and the proper classification of childhood disorders. A major difficulty in answering these questions stems from researchers' failure to agree on criteria for what constitutes psychiatric disturbance of childhood. Without such criteria, they can agree on neither rates of *appearance* nor rates of *disappearance* of disorders. As the studies presented indicate, being seen by a psychiatrist cannot be a satisfactory measure of the number of children with psychiatric disorders, since that figure can vary between 2 per cent of Swedish children seen in a ten-year period to more than ten times that rate, 21 per cent referred to a psychiatrist before the age of 7, when parents are particularly accepting of psychiatric intervention.

Nor can the presence of at least one symptom be a satisfactory measure, since all children would qualify at one age or another (33). At the present time, we can say only that difficulties severe enough to cause concern to teachers and psychiatrists occur in at least 7 per cent of children (the proportion diagnosed as psychiatrically ill in a ten-year period in one Swedish study). But the figure may be considerably greater than 7 per cent, since another Swedish study found 25 per cent in need of treatment.

We do know that most children identified as having problems in clinics and in schools are considered as well or improved a year or two later, but no study has yet attempted to ascertain the length of time between onset and disappearance either for individual symptoms or for syndromes. Consequently, we cannot yet prognosticate with any assurance how long a child's disturbance is likely to last. We know only that neurotic symptoms are much more likely to dissipate or to improve in a short time than are symptoms of psychosis or antisocial behaviour.

Nor have follow-up studies yet produced a comprehensive classificatory system for childhood disorders. To develop a comprehensive classification will require both more 'backward' follow-ups of representative samples of adults with well-defined syndromes and more prospective follow-up studies into adulthood of children identified by symptom clusters and family histories. Only when the whole spectrum of adult disorders has been identified as beginning or not beginning in childhood, and when these early appearing adult disorders can be separated from disorders limited to childhood, can we begin to develop a coherent classification scheme for childhood disorders. The studies presented here have made only a small beginning in this enterprise. They have shown that psychosis appearing before puberty is never or almost never an early form of adult schizophrenia, but that the psychotic child does have an illness that continues into adulthood, where it accounts for only a small proportion of hospitalized adults. They have shown that hysteria can, but rarely does, begin in childhood, while the more common adult neuroses seem *not* to begin in childhood. Finally, they have shown that antisocial personality begins almost exclusively in childhood. We as yet do not know what proportion of childhood disorders these early forms of adult disorders account for, nor can we readily distinguish by symptom pattern childhood disorders which will and which will not be continued into adulthood. Indeed our best clue at the present time as to whether or not a disorder will be transient seems to be the severity of incapacitation and number of symptoms rather than its classification by syndrome.

We are on firmer ground at the present time when it comes to predicting *which* children will develop disorders, however we define them. We know that psychiatric illness in the parents is reflected in an elevated rate of problems in their children, and that antisocial parents have antisocial children. When the parents' antisocial behaviour takes the form of physical abuse of the child, the child shows very severe disability. We also know that childhood disorder is more common in the lower classes and that this relationship may be largely accounted for by the higher prevalence of antisocial and psychiatrically ill parents in the lower classes. To what extent the influence of the parent's disturbance on the child's behaviour is genetic and to what extent it is environmental is still an unanswered question.

We can also predict certain disorders of later childhood from behaviour first observable around the age of starting school: a combination of poor relationships with schoolmates, working below intellectual capacity, and failing to accept the teacher's authority predict delinquency. A high level of neurotic symptoms predicts poor school success. Early behavioural predictors of other kinds of childhood disorders are not yet clearly identified.

We also know a good deal about the adult prognosis for disturbed children. The weight of evidence at the present time is that two kinds of childhood disorders, psychosis and serious antisocial behaviour, have gloomy prognoses for adult life. Psychosis in children, whether it begins before puberty or whether it is schizophrenia beginning in adolescence, almost always persists into adulthood. Most psychotic children will spend much of their lives in institutions. Serious antisocial behaviour in childhood is not so uniformly disastrous, but in a large minority of cases it presages life-long problems with the law, inability to earn a living, defective interpersonal relationships, and severe personal distress. In fact, if one could successfully treat the antisocial behaviour of childhood, the problems of adult crime, alcoholism, divorce, and chronic unemployment might be vastly diminished. These two disorders, psychosis and antisocial behaviour, along with severe mental retardation, which we have not discussed here, are the childhood psychiatric categories for which serious adult difficulties can be predicted. Neurotic disorders of childhood, quite to the contrary, have very good long-term prognoses, no matter how disturbing they may be to the family and child while they exist.

At the present time, unfortunately, there is little substantive evidence that childhood psychosis or antisocial behaviour can be either prevented or treated successfully. It is, however, encouraging that follow-up methods provide us with a technique for evaluating our attempts at prevention and treatment that produces clear and consistent answers. When we do find successful techniques for preventing or modifying childhood disorders, we can expect follow-up studies to demonstrate their *effectiveness* as dramatically and consistently as they now demonstrate our present failures in prevention and cure.

REFERENCES

1. ANNESLEY, P. T. (1961). 'Psychiatric illness in adolescence: presentation and prognosis', *J. ment. Sci.* **107**, 268–78.
2. BENDER, L. (1953). 'Childhood schizophrenia', *Psychiat. Quart.* **27**, 1–19.
3. BOWER, E. M., SHELLHAMER, T. A., and DAILY, J. M. (1960). 'School characteristics of male adolescents who later become schizophrenic', *Amer. J. Orthopsychiat.* **30**, 712–29.
4. BOWLBY, J., AINSWORTH, M., BOSTON, M., and ROSENBLUTH, D. (1956). 'The effects of mother–child separation: a follow-up study', *Brit. J. med. Psychol.* **29**, 211–44.
5. BURT, C. (1925). *The Young Delinquent* (New York: D. Appleton and Company).
6. CANAVAN, M., and CLARK, R. (1923). 'The mental health of 463 children from dementia praecox stock, I', *Ment. Hyg.* (*N.Y.*), **7**, 137–48.
7. —— —— (1923). 'The mental health of 581 offspring of non-psychotic parents', ibid. **7**, 770–8.
8. CONGER, J. J., and MILLER, W. C. (1966). *Personality, Social Class, and Delinquency* (New York: J. Wiley).
9. CRAIG, M., and GLICK, S. J. (1963). 'Ten years experience with the Glueck Social Prediction Table', *Crime and Delinquency*, **9**, 249–61.
10. CUMMING, J., and MCCAFFREY, I., (1964). 'Persistence of emotional disturbances reported among second and fourth grade children', *Onondaga County School Studies. Interim Report No. 1* (Mental Health Research Unit, New York State Department of Mental Hygiene, Syracuse, New York).

11. Cunningham, J. M., Westerman, H. H., and Fischoff, J. (1956). 'A follow-up study of patients seen in a psychiatric clinic for children', *Amer. J. Orthopsychiat.* **26**, 602–12.
12. Douglas, J. W. B. (1964). *The Home and the School* (London: Macgibbon and Kee).
13. —— (1966). 'The school progress of nervous and troublesome children', *Brit. J. Psychiat.* **112**, 1115–16.
14. Eaton, L., and Menolascino, F. L. (1967). 'Psychotic reactions of childhood: a follow-up study', *Amer. J. Orthopsychiat.* **37**, 521–9.
15. Elmer, E., and Gregg, G. S. (1967). 'Developmental characteristics of abused children', *Pediatrics*, **40**, part i, 596–602.
16. Frazee, H. E. (1953). 'Children who later became schizophrenic', *Smith College Studies in Social Work*, **23**, 125–49.
17. Gildea, M. C.-L., Glidewell, J. C., and Kantor, M. B. (1967). 'The St. Louis School Mental Health Project: history and evaluation', in *Emergent Approaches to Mental Health Problems*, eds. Cowen, E. L., Gardner, E. A., and Zax, M. (New York: Meredith Publishing Company).
18. Glavin, J. P. (1967). ' "Spontaneous" improvement in emotionally disturbed children', Doctoral dissertation (George Peabody College for Teachers).
19. Glueck, S., and Glueck, E. (1940). *Juvenile Delinquents Grown Up* (New York: The Commonwealth Fund).
20. —— —— (1950). *Unraveling Juvenile Delinquency* (New York: The Commonwealth Fund).
21. Goldberg, E. M., and Morrison, S. L. (1963). 'Schizophrenia and social class', *Brit. J. Psychiat.* **109**, 758–802.
22. Hagnell, O. (1966). *A Prospective Study of the Incidence of Mental Disorder* (Stockholm: Svenska Bokförlaget).
23. Havighurst, R. J., Bowman, P. H., Liddle, G. P., Matthews, C. V., and Pierce, J. V. (1962). *Growing Up in River City* (New York: J. Wiley).
24. Healy, W., and Bronner, A. F. (1936). *New Light on Delinquency and its Treatment* (New Haven: Yale University Press).
25. Heston, L. L. (1966). 'Psychiatric disorders in foster home reared children of schizophrenic mothers', *Brit. J. Psychiat.* **112**, 819–25.
26. Huffman, P. W., and Wenig, P. W. (1954). 'Prodromal behaviour patterns in mental illness' (Department of Public Welfare, State of Illinois, mimeo.).
27. Jonsson, G., and Kalvesten, A.-L. (1964). *222 Stockholmspojkar, Uppsala.* A summary in English appears in Gustav Jonsson (1967) 'Delinquent boys, their parents and grandparents', *Acta psychiat. scand.*, Suppl. 195, Appendix I, 227–56.
28. Kanner, L., and Eisenberg, L. (1955). 'Notes on the follow-up studies of autistic children', in *Psychopathology of Children*, 227–39 (Grune and Stratton, Inc.).
29. Lane, E. A., and Albee, G. W. (1965). 'Childhood intellectual differences between schizophrenic adults and their siblings', *Amer. J. Orthopsychiat.* **35**, 747–53.
30. —— —— (1966). 'Comparative birth weights of schizophrenics and their siblings', *J. Psychol.* **64**, 227–31.

31. LEWIS, A. J. (1957). 'The offspring of parents both mentally ill', *Acta genet.* **7**, 349–65.

32. LEWIS, W. W. (1965). 'Continuity and intervention in emotional disturbance: a review', *Exceptional Children*, **31**, 465–75.

33. MACFARLANE, J. W., ALLEN, L., and HONZIK, M. P. (1954). *A Developmental Study of the Behavior Problems of Norman Children between 21 months and 14 years* (Berkeley and Los Angeles: University of California Press).

34. MASTERSON, J. F. Jr. (1958). 'Prognosis in adolescent disorders', *Amer. J. Psychiat.* **114**, 1097–1103.

35. —— (1967). 'The symptomatic adolescent five years later: he didn't grow out of it', ibid. **123**, 1338–45.

36. MCCORD, W., and MCCORD, J. (1959). *Origins of Crime* (New York: Columbia University Press).

37. MEYER, H. J., BORGATTA, E. F., and JONES, W. C. (1965). *Girls at Vocational High* (New York: Russell Sage Foundation).

38. MICHAEL, C. M. (1957). 'Relative incidence of criminal behavior in long-term follow-up studies of shy children', *Dallas Medical Journal*, January.

39. —— MORRIS, D. P., and SOROKER, E. (1957). 'Follow-up studies of shy withdrawn children: II. Relative incidence of schizophrenia', *Amer. J. Orthopsychiat.* **27**, 331–7.

40. MOLOF, M. J. (1967). 'Forestry Camp Study: Comparison of recidivism rates of camp-eligible boys randomly assigned to camp and to institutional programmes', *Research Report No. 53* (State of California, Department of the Youth Authority).

41. MORRIS, H. H. Jr., ESCOLL, P. J., and WEXLER, R. (1956). 'Aggressive behavior disorders of childhood: a follow-up study', *Amer. J. Psychiat.* **112**, 991–7.

42. MULLIGAN, G., DOUGLAS, J. W. B., HAMMOND, W. A., and TIZARD, J. (1963). 'Delinquency and symptoms of maladjustment: the findings of a longitudinal study', *Proc. roy. Soc. Med.* **56**, 1083–8.

43. OTTERSTRÖM, E. (1946). 'Delinquency and children from bad homes', *Acta paediat. (Uppsala)*, **33**, Suppl. 5.

44. POWERS, E., and WITMER, H. (1951). *An Experiment in the Prevention of Delinquency* (New York: Columbia University Press).

45. PRITCHARD, M., and GRAHAM, P. (1966). 'An investigation of a group of patients who have attended both the child and adult departments of the same psychiatric hospital', *Brit. J. Psychiat.* **112**, 603–12.

46. ROBERTS, C. F., Jr. (1967). 'A follow-up study of the juvenile drug offender' (Institute for the Study of Crime and Delinquency, Sacramento, California, October).

47. ROBINS, E., and O'NEAL, O. (1953). 'Clinical features of hysteria in children, with a note on prognosis', *The Nervous Child*, **10**, 246–71.

48. ROBINS, L. N. (1966). *Deviant Children Grown Up* (Baltimore: The Williams and Wilkins Company).

49. —— and LEWIS, R. G. (1966). 'The role of the antisocial family in school completion and delinquency: a three-generation study', *Sociological Quarterly*, **7**, 500–14.

50. —— (1968). 'Negro homicide victims—who will they be?', *Trans-action*, **5**, 15–19.

51. —— Murphy, G. E., and Breckenridge, M. B. (1968). 'Drinking behavior in young urban Negro men', *Quart. J. Stud. Alcohol*, **29**, 657–84.

52. Roff, M. (1956). 'Preservice personality problems and subsequent adjustments to military service: gross outcome in relation to acceptance–rejection at induction and military service', *Report 55–138* (School of Aviation Medicine, USAF, Randolph AFB, Texas).

53. —— (1961) 'The service-related experience of a sample of juvenile delinquents', *Report No. 61–1* (U.S. Army Medical Research and Development Command).

54. —— (1961). 'Childhood social interactions and young adult bad conduct', *J. abnorm. soc. Psychol.* **63**, 333–7.

55. —— (1963). 'The service-related experience of a sample of juvenile delinquents: II. A replication on a larger sample in another state', *Report No. 63–2* (U.S. Army Medical Research and Development Command).

56. Rutter, M., Birch, H. G., Thomas, A. *et al.* (1964). 'Temperamental characteristics in infancy and the later development of behavioural disorders', *Brit. J. Psychiat.* **110**, 651–61.

57. —— (1965). 'The influence of organic and emotional factors on the origins, nature and outcome of childhood psychosis', *Developmental Medicine and Child Neurology*, **7**, 518–27.

58. —— and Lockyer, L. (1967). 'A five to fifteen year follow-up study of infantile psychosis: I. Description of sample', *Brit. J. Psychiat.* **113**, 1169–82.

59. —— Greenfeld, D., and Lockyer, L. (1967). 'A five to fifteen year follow-up study of infantile psychosis: II. Social and behavioural outcome', ibid. **113**, 1183–99.

60. Shepherd, M., Oppenheim, A. N., and Mitchell, S. (1966). 'Childhood behaviour disorders and the child guidance clinic: an epidemiological study', *J. Child Psychol. Psychiat.*, **7**, 39–52.

61. Shirly, M., Baum, B., and Polsky, S. (1940–1). 'Outgrowing childhood's problems: a follow-up study of child guidance patients', *Smith College Studies in Social Work, II*, 31–60.

62. Tait, C. D., Jr., and Hodges, E. F., Jr. (1962). *Delinquents, their Families, and the Community* (Springfield, Illinois: Charles C. Thomas).

63. Thomas, A., Chess, S., and Birch, H. G. (1968). *Temperament and Behavior Disorders in Children* (New York: New York University Press).

64. Warren, W. (1965). 'A study of adolescent psychiatric in-patients and the outcome six or more years later: II. The follow-up study', *J. Child Psychol. Psychiat.* **6**, 141–60.

65. Westman, J. C., Rice, D. L., and Bermann, E. (1967). 'Nursery school behaviour and later school adjustment', *Amer. J. Orthopsychiat.* **37**, 725–31.

66. Wilner, D. M., Walkley, R. P., Pinkerton, T. C., and Tayback, M. (1962). *The Housing Environments and Family Life* (Baltimore: Johns Hopkins Press).

67. World Health Organization (1962). 'Deprivation of maternal care: a reassessment of its effects', *Public Health Papers, No. 14*, 7–165.

Discussion

I

MICHAEL L. RUTTER
Institute of Psychiatry, University of London

Dr Lee Robins has given a most thoughtful and masterly review which leaves little room for disagreement. There are only two conclusions with which I would like to take issue: that on the advantages of the follow-up study for sampling and that on studies of treatment. There is also one important parameter which seems to have been ignored—sex differences.

SAMPLING

Dr Robins concludes that follow-up studies allow a more complete sampling of the population than can be achieved by cross-sectional studies. In support, she gives the example of a cross-sectional study of 7th graders which is undertaken to find out how many children develop behaviour problems before puberty. As she points out, this would leave out children who never entered 7th grade—those in hospitals, correctional institutions, and the like. In contrast a representative sample of very young children which was followed to age 12 years would show how many children were not in ordinary schools. But surely this is merely comparing a bad cross-sectional study with a good follow-up study. There is no reason at all why cross-sectional studies should not encompass the whole population of children—not just those at school. We have done this, for example, in our surveys on the Isle of Wight (53, 54, 55, 56). In fact, in follow-up studies it is much more difficult to ensure that all children are traced. In Dr Robins's own study she had an unparalleled level of success in tracing children, but even so, personal interviews were obtained with only 76 per cent of the subjects. It would have to be a rather incompetent cross-sectional study not to do better than that.

I would rather conclude that longitudinal studies should *never* be done to obtain cross-sectional data. It is an expensive and inefficient method for this purpose. On the other hand, where longitudinal information is required there is no adequate substitute for a longitudinal approach. Retrospective information from a cross-sectional study is open to far too many biases for this to be an alternative, except at a preliminary stage.

This is particularly the case when deaths may affect the issue. An illustration is provided by prospective and retrospective studies of coronary heart disease (32). It seems that the psychological characteristics of men who *die* from myocardial infarction differ in some respects from those who *survive*. A retrospective study of the psychological antecedents of myocardial infarction would necessarily have entirely missed this.

However, it is important to note that follow-up studies also have their disadvantages in any examination of the association between antecedent factors and later outcome. For example, Dr Robins's follow-up study (46) of children seen at a child-guidance clinic showed that a high proportion of those with antisocial behaviour became sociopaths as adults. She concluded that sociopathic personality is a psychiatric disease which begins in early childhood and continues into adult life. However, the follow-up in itself does not tell us how many adult sociopaths did *not* show antisocial behaviour in childhood.

Some estimate may be obtained from the control group, where 2 per cent became sociopaths compared with 28 per cent of the children referred to a child-guidance clinic for antisocial behaviour—an impressive difference, but the strength of the association is also dependent on the relative frequency in the general population of 'controls' and of children referred to a child-guidance clinic for antisocial behaviour. This raises problems, in that we have no means of knowing from these data how antisocial children referred to a psychiatrist differ from those not referred. Also Dr Robins does not give findings separately for boys and girls—a point to which I shall return. However, let us assume for the moment that antisocial children attending clinics do not differ from those not attending clinics and let us assume a rate of antisocial disorder (of the type referred to clinics) of 5 per cent. Given a population of 1,000 there would be 50 antisocial children, of whom 14 would become sociopaths. Of the remaining 950 children, 19 would become sociopaths. In other words most sociopaths would *not* have shown antisocial behaviour in childhood. If the Robins findings apply only to antisocial children attending clinics and not to those who do not get referred to psychiatrists, the difference would be even more striking—in that case nearly all sociopaths would come from the control group.

Of course in this calculation I have had to make all sorts of assumptions which might prove to be false. I certainly do not intend to criticize the Robins follow-up, which undoubtedly is by far and away the finest of its kind. My point is simply that to determine the strength of an association a combination of longitudinal and cross-sectional studies is required. Thus, in order to tackle this problem in relation to adolescent disorders we have used this combined approach in the Isle of Wight surveys. We carried out a total population survey of 10-year-olds in 1964–5 and now we are both following up the children who then showed disorder and also resurveying the total population of 14-year-olds. By this means we can compare those disorders in 14-year-olds which represent a continuation of the problems seen at age 10 years with those disorders which have arisen anew in children who were normal four years ago.

STUDIES OF TREATMENT

The paucity of adequate studies of the efficacy of psychiatric treatment in childhood is deplorable and I entirely support Dr Robins's criticisms in this connection. Also I am sure she is right to question whether it is reasonable to evaluate treatment by studying improvement rates in a total clinic population. In other branches of medicine no one would dream of examining the effects of treatment in a study which specified neither the type of disorder nor the type of treatment. It would be pointless to do so. Yet this is exactly what many studies of treatment of child psychiatric disorders have done. I find it all the more curious that after strongly criticizing this approach, Dr Robins finishes the section of the paper devoted to the treatment by quoting the Buckinghamshire study (59) as the one rigorous study of the effects of treatment on *non*-delinquent and *non*-psychotic children. This is certainly one of the best surveys of child disorders but, in my view, it is quite inappropriate for any study of treatment.

It falls into the very category of studies which Dr Robins had condemned a few pages earlier. It assesses the effects of therapy of an unknown type and unknown quality on a group of children with disorders of largely unknown diagnosis. The groups were ill-matched, in that the clinic group included significantly fewer children with mild disorders[1] (48 per cent in the controls and 26 per cent in the cases) and significantly more from homes where the father was absent and the mother mentally ill. Furthermore, the follow-up period was two years, too long to assess the effects of treatment on a disorder with a high remission rate. In these circumstances treatment is to be assessed, not in relation to 'cure', but rather in terms of its power to shorten the duration of disorder.

By restricting attention to studies where a long follow-up has given 'an opportunity for an important amount of change to occur', Dr Robins has ruled out of court studies showing that treatment has brought about a marked change in a short period of time. Having

[1] The χ^2 in the paper was not significant owing to the separation of moderate and severe disorders. However, the severe category was very small and if the χ^2 is repeated after combining these groups the difference is significant.

said that, I must admit that such studies of treatment are few and far between. Nevertheless there is some modest evidence that short-term psychotherapy (18), drug treatment (1, 12, 14, 18, 69), and conditioning techniques (3, 25, 33) can, within limits, be effective forms of treatment for child psychiatric disorders if correctly employed for the right type of patient. However, follow-up studies in relation to both drug treatment (58) and conditioning techniques (34) have sometimes shown relapse rates to be alarmingly high. We are still at a very early stage in attempts to evaluate treatment (16, 28) and many studies have been distressingly naïve and poorly controlled (62).

Nevertheless, I think it would be wrong to conclude that no progress has been made and we must not fall into the error of equating equivocal results with a proof of the failure of treatment. In this respect, Dr Robins seems to fall momentarily from her own high standards in concluding on childhood psychosis that '. . . all carefully controlled studies have agreed in showing poor results no matter what techniques have been attempted'. What carefully controlled studies? There are several follow-up studies of psychotic children, but none of these was a controlled trial of any treatment. My own study (55) suggested that schooling *did* influence the outcome, but this finding must be treated with reserve, in that children were not randomly allocated to schooling. Prospective studies of treatment are much needed. It is rarely practical to compare treatment with no treatment. Indeed, Eisenberg's studies (19, 39) suggest that if children know they are having no treatment this may have a significantly deleterious effect. It is probably more useful to compare different forms of treatment. We are currently attempting to do this by comparing the progress of autistic children in three units using different therapeutic methods—one having a strongly educational skill-oriented approach, one using psychotherapeutic techniques based on a one-to-one relationship, and the third using an intermediate approach. We are also just about to start a study comparing the effects of contingent and non-contingent reinforcement in speech and language training for autistic children.

SEX DIFFERENCES

The third aspect on which I want to comment is that of sex differences. From Dr Robins's review one would never guess that boys and girls are different.

Yet her own follow-up study showed clear sex differences (46). Most children of both sexes (74 per cent of boys and 68 per cent of girls) had been referred for antisocial problems, there being no appreciable sex difference in this. But in adult life only 10 per cent of the boys compared with 44 per cent of the girls were neurotic. In boys antisocial behaviour in childhood led to sociopathy in the adult, in girls it more often led to neurosis, particularly hysteria. There was a similar tendency in the Pritchard and Graham (44) study but the numbers were too small to assess the significance of the sex difference.

Other follow-up studies have also shown sex differences in the associations between child psychiatric disorder and adult mental illness. For example, the study of children seen at the Judge Baker Clinic who later developed schizophrenia (66) showed that certain neurotic symptoms were antecedents of schizophrenia in boys but not in girls (22). On the other hand maternal psychopathology was associated with schizophrenia in girls but not in boys (23). The criteria for the diagnosis of schizophrenia were not made explicit and it is uncertain how wide a group of conditions were included.

In this context, it may be relevant that the concordance rates for schizophrenia in adult twins appear to be higher for women than for men (31, 47). This may prove to be an artefact of sampling and in any case, as the sex difference applies to both monozygotic and dizygotic twins, it is likely to be an environmental rather than a genetic effect (60). Nevertheless, sex differences will have to be looked for in any further studies of schizophrenia, such as the studies of adoptive and biological parents of schizophrenics being carried out by Kety, Rosenthal, Wender, and Schulsinger (30, 48, 70) and of the children of schizophrenic mothers in Meduick's and Schulsinger's study (38). Incidentally, these are, I think, first-rate follow-up studies which should have found a place in Dr Robins's review. The first group of studies offers much the most convincing evidence to date for a genetic factor in schizophrenia. As MacMahon (37) has commented, by pointing to the concordance rates in tuberculosis and poliomyelitis, proof of a genetic influence does not rule out the possibility of there being a decisive environmental effect as well, but still the findings demonstrate that genetic influences are not just a factor, but a *significant determinant* of schizophrenia (17).

Not only do there seem to be sex differences in abnormal development, but also there are sex differences in normal personality development. Both American and British studies (6, 7, 27, 29, 40, 41) suggest that different factors are predictive of psychological development in boys and in girls.

To illustrate further the importance of sex differences I would like to turn to a study we have been carrying out into the effects on children of mental illness in one parent.[1] The study is following

---

[1] This investigation is being carried out in conjunction with Dr Philip Graham, Mr Quinton, Mr Ziffo, Mrs George, Miss Rowlands, Miss Tupling, and Miss Osborn.

families over a four-year period but the data I am going to present now refer entirely to the findings in the first year.

The sample consists of the families of all patients living in the old Borough of Camberwell in London who spoke colloquial English at home, who had one or more children under the age of 15 years, and who were *newly* referred to a psychiatric clinic during the course of a ten-month period—some 200 families in all. Two-fifths of these were intensively investigated and three-fifths (chosen at random) were seen once only for a more limited assessment. The patients were identified shortly after the time of their first psychiatric attendance by means of the Camberwell psychiatric register (71); they attended some dozen different hospitals. The original ten-month sample contained fewer men than women and there were few psychotics, so an additional sample of another 50 or so families were added, consisting of consecutive referrals in these categories. For present purposes the samples have been combined.

The refusal rate in the study was low—under 5 per cent. Both the husband and wife were seen separately by different interviewers for interviews of 2–3 hours, sometimes extending to two sessions. Information was systematically obtained on the patient's disorder, the social context of their symptoms and their impact on the family, the health and behaviour of the spouse and children, many aspects of family life and activities (including contacts with kin and friends, leisure pursuits, and the distribution of household tasks and child-care activities), the nature and extent of parent–child interaction, and the relationships between different members of the family.

Some information was obtained on all children, but detailed information was got for only two children in each family (chosen at random). Information on the children's behaviour at school was obtained by means of a questionnaire completed by teachers. Similar questionnaires were also completed for two other children in the same class (matched for age and sex). This was done in such a way that the teacher did not know who were the study children and who were the controls.

Considerable attention was paid to the standardization of interviewing methods, which were developed in conjunction with Dr George Brown, and on attaining high reliability and validity. In this we were reasonably successful (8, 52).

For present purposes I am going to use the teachers' questionnaire as the only measure of the children's psychiatric state but with only a very few exceptions the results are broadly similar using other means of assessment. The reason for using the teachers' questionnaire is that it provides the harshest test of associations between the family situation and child psychiatric disorder, our main concern (49). On the questionnaire, the children's behaviour is being assessed outside the home by someone unacquainted with details of the child's family life. As the questionnaires were completed by teachers and the family measures made by us, there is no serious possibility of contamination of results. Any association found using this approach can be assumed to represent real relationships of an important kind.

The questionnaire has been shown in previous studies to be a reasonable index of psychiatric disorder (50, 54), although like any questionnaire it is rather a crude index. As well as providing an over-all assessment of the presence or absence of psychiatric disorder, it can also be scored to differentiate neurotic disorders from antisocial disorders.

Two types of comparison can be made: (1) between children in the patient's family and children in the general population, in other words case versus control comparisons, and (2) within sample comparisons. In this case we will be comparing children all of whom have a parent under psychiatric care, but subdividing them in some way to differentiate them according to the diagnosis of the parent or the quality of the parental marriage or some other measure of this kind. In each case, the concern will be to relate family characteristics to the behaviour of the children.

Psychiatric disorder was over twice as common in the daughters of patients as in the control girls (30–2 per cent versus 8–15 per cent), regardless of whether the patient was the father or mother. In sharp contrast, the rate of disorder in the sons of patients was only slightly (and non-significantly) above the rate in the control boys. In the

TABLE 2.1. *Teacher questionnaire findings for cases and controls (matched for age, sex, and school class)*

	Male patients				*Female patients*			
	Boys		*Girls*		*Boys*		*Girls*	
	No.	%	*No.*	%	*No.*	%	*No.*	%
Cases								
Normal on Q	51		50		53		55	
Deviant on Q	24	(32)	24	(32)	29	(35)	24	(30)
Total	75		74		82		79	
Controls								
Normal on Q	106		123		120		145	
Deviant on Q	44	(29)	22	(15)	43	(26)	13	(8)
Total	150		145		163		158	

controls, as in virtually all previous studies (54), the rate of disorder was twice as high in boys as in girls. In the cases, on the other hand, disorder was equally common in the two sexes (Table 2.1).

TABLE 2.2. *Age differences for girls on teacher questionnaire*

	Age of child: months					
	60–95		*96–143*		*144–80*	
	No.	%	*No.*	%	*No.*	%
Cases						
Deviant	16	(33)	18	(26)	14	(39)
Total	49		68		36	
Controls						
Deviant	10	(10)	16	(12)	8	(11)
Total	97		135		71	

The case-control difference applies at all ages for girls (Table 2.2) but at no age is there a significant difference for boys (Table 2.3). Note also that for boys, disorders show a significant age trend in both the cases and controls—disorders being commoner in younger boys. In contrast, in neither cases nor controls is there an age trend for girls.

TABLE 2.3. *Age differences for boys on teacher questionnaire*

	Age of child: months					
	60–95		*96–143*		*144–80*	
	No.	%	*No.*	%	*No.*	%
Cases						
Deviant	23	(48)	20	(32)	10	(21)
Total	48		62		47	
Controls						
Deviant	37	(39)	33	(27)	17	(18)
Total	95		124		94	

Table 2.4 shows that the increased rate of disorder in the cases is due more to an increase in antisocial disorder than in neurotic disorder. This tendency is present in the girls and is more definite in the

TABLE 2.4. *Diagnosis on teacher questionnaire for cases and controls*

Questionnaire diagnosis	*Boys*				*Girls*			
	Cases		*Controls*		*Cases*		*Controls*	
	No.	%	*No.*	%	*No.*	%	*No.*	%
Normal	104		226		105		268	
Deviant	53		87		48		36	
Neurotic	12	(8)	38	(12)	20	(13)	17	(6)
Antisocial	33	(21)	43	(14)	26	(17)	15	(5)
Mixed	8	(5)	6	(2)	2	(1)	4	(1)
Total	157		313		153		304	

boys. If the 'mixed' disorders are combined with the antisocial disorders (and other work suggests that they have more in common with antisocial than with neurotic disorders), the case-control difference is significant for antisocial disorder in boys whereas there is no increase in neurotic disorder.

The diagnosis of the parental illness appeared irrelevant. In all diagnostic groups there were case-control differences for girls, but not for boys. However, there was a tendency for there to be case-control differences in boys when the father had an antisocial personality. In girls the presence or absence of a personality disorder in the parent was without effect—with or without personality disorder in the parent there were large case-control differences.

So far, we have been concerned solely with case-control comparisons. Here the differences were largely restricted to girls. If we turn now to the within-sample comparisons the reverse is found. In nearly all cases the differences were confined to boys.

TABLE 2.5. *Marriage rating and teacher questionnaire designation of boys*

Marriage rating	*Teacher questionnaire*						
	Neurotic		*Antisocial*		*Total deviant*		
	No.	%	*No.*	%	*No.*	%	*Total no.*
'Good'	1	(4)	0	(0)	1	(4)	23
'Fair/poor'	5	(7)	11	(15)	21	(30)	71
'Very poor'	6	(10)	27	(46)	37	(63)	59

For example, Tables 2.5 and 2.6 show disorder in relation to the rating of the parental marriage. This is a highly reliable rating based on many factors including the frequency of quarrelling, the amount

TABLE 2.6. *Marriage rating and teacher questionnaire designation of girls*

	Teacher questionnaire						
	Neurotic		*Antisocial*		*Total deviant*		
Marriage rating	*No.*	*%*	*No.*	*%*	*No.*	*%*	*Total no.*
'Good'	2	(13)	2	(13)	4	(27)	15
'Fair/poor'	9	(12)	12	(16)	23	(31)	74
'Very poor'	7	(14)	10	(20)	17	(33)	51

of warmth between husband and wife, the hostility and criticism expressed by one about the other, the amount of dissatisfaction expressed about the marriage, and the quality of husband–wife interaction in household activities and leisure pursuits. The marriage rating bore *no* relation to the rate of disorder in girls, but in boys the worse the parental marriage the higher the rate of disorder in the sons. However, the marriage rating was related only to antisocial disorder in sons and not to neurotic disorder.

It can be seen from Table 2.7 that this association applies whether or not the parent had a personality disorder, but the rate of disorder was most high when there was *both* a parental personality disorder *and* a poor marriage.

TABLE 2.7. *Personality of patient, marriage rating, and boys' teacher questionnaire designation*

	Marriage rating								
	Good			*Fair/poor*			*Very poor*		
Personality of patient	*No. deviant*	*%*	*Total no.*	*No. deviant*	*%*	*Total no.*	*No. deviant*	*%*	*Total no.*
Abnormal	1	(14)	7	7	(29)	24	31	(76)	41
Normal	0	(0)	16	13	(29)	45	6	(33)	18

The same findings hold when different aspects of the marriage situation, such as the frequency of quarrelling, are considered. It also holds when the parental situation is considered. 'Broken homes' were related to antisocial disorder in boys but not in girls. The same applies to admission to children's home or to foster care. Similar findings emerge also from a consideration of parent–child interaction. Thus, discord and disruption in the home were consistently associated with antisocial disorder in boys but not in girls. No consistent associations were found between family characteristics and neurosis in either boys or girls.

TABLE 2.8. *Parental quarrels and teacher questionnaire designation of boys*

Number of quarrels (in 3 months)	*Teacher questionnaire* *Neurotic* *No.*	%	*Antisocial* *No.*	%	*Total deviant* *No.*	%	*Total no.*
0–2	0	(0)	3	(10)	4	(13)	31
3–9	1	(6)	2	(13)	3	(19)	16
10 or more	6	(16)	15	(43)	22	(59)	37

TABLE 2.9. *Parental quarrels and teacher questionnaire designation of girls*

Number of quarrels (in 3 months)	*Teacher questionnaire* *Neurotic* *No.*	%	*Antisocial* *No.*	%	*Total deviant* *No.*	%	*Total no.*
0–2	3	(9)	5	(15)	8	(24)	34
3–9	0	(0)	0	(0)	0	(0)	10
10 or more	4	(12)	7	(21)	13	(38)	34

The association between family discord and antisocial behaviour is a strong one, but which leads to which and how it does so needs also to be considered. As Bell (4) pointed out in a thoughtful paper in 1968, parent–child correlations may represent the effects of children on their parents rather than the other way round. Nevertheless in the present case it seems probable that the main interaction is in the direction of parent to child, if only because the method of sampling was deliberately chosen so that in most cases the parental disorder antedated the child's disorder.

This is in keeping with other prospective studies which have all shown family discord to be a powerful predictor of later disturbance in the children (13, 15, 21, 36, 43). None of these studies examined sex differences. However, it is very striking that studies of mortality in spouses and children following bereavement have shown that there is a sharply increased mortality among bereaved close relatives compared with controls and that this difference is significantly greater for males than for females (45).

However, if we conclude that males are more susceptible than females to the adverse effects of family discord and disruption, this still leaves the problem of why the sex difference is the reverse of that found when considering case-control comparisons. Elsewhere (51) I have considered the possible explanations for this curious finding. I concluded that it is unlikely that methodological biases are responsible. Rather it seems that either the differences *between* cases and controls are different in *quality* from those *within* cases and/or that different mechanisms are involved for boys and for girls. It could be that boys and girls are susceptible to different types of family stress.

TABLE 2.10. *Parental situation and teacher questionnaire designation of boys*

	Teacher questionnaire						
	Neurotic		*Antisocial*		*Total deviant*		
Parental situation	*No.*	*%*	*No.*	*%*	*No.*	*%*	*Total no.*
Child living with both natural parents	8	(6)	24	(19)	33	(26)	125
Other situations	4	(9)	19	(41)	28	(61)	46

TABLE 2.11. *Parental situation and teacher questionnaire designation of girls*

	Teacher questionnaire						
	Neurotic		*Antisocial*		*Total deviant*		
Parental situation	*No.*	*%*	*No.*	*%*	*No.*	*%*	*Total no.*
Child living with both natural parents	16	(13)	20	(16)	38	(30)	126
Other situations	4	(13)	7	(23)	11	(35)	31

This could certainly be the case in that there are important sex differences in parent–child interaction from an early age in non-human primates as well as in humans (51). Also the same family situation might constitute a different stimulus for girls than it does for boys. For example, the same family situation might constitute a greater stress to the child who has the greater contact with the disturbed parent or who has to take over more of the home responsibilities of the sick parent.

So far we have only just begun to explore these issues and it is too early to say how important or unimportant they may be. In the meantime, however, let me suggest another explanation. Perhaps the effect in girls is largely genetic and in the boys largely psychosocial in origin. The hypothesis here runs as follows. First, as Robins (46) showed in her follow-up study, there is very little continuity between child neurosis and adult neurosis. Most neuroses in childhood are short-lived and most adult neuroses arise in adult life, not in childhood. But, as Pritchard and Graham (44) found, some child neuroses do persist into adult life. Apart from certain types of phobia, as shown by Marks and Gelder (35), we do not yet know how to distinguish the child neuroses which are confined to childhood from those neuroses which show continuity between childhood and adult life and those neuroses which start for the first time in adulthood.

Secondly, in adults, neuroses and depression are much commoner in women than in men, whereas in children the sex difference is much less marked.

Lastly, in this connection, most of the parents in the present study had neurosis or depression. If the increased rate of disorder in the children of patients was genetically determined, then this should increase the rate in girls more than in boys—because adult neurosis and depression are particularly associated with the female sex. We are here assuming that this sex difference is genetically determined.

There is not too much evidence in support of this admittedly speculative hypothesis, but it may be noted that apart from the children of mentally ill parents the only other situation where the rate of disorder is as high in girls as in boys is when the children have some neuro-epileptic disorder (42, 54, 56). This suggests a biologic effect on susceptibility.

It is appreciated that this hypothesis implies that antisocial disorder in girls may be genetically associated with adult neurosis. Unlikely though this seems at first sight, it will be recalled that the Robins (46) follow-up offers some support for this notion, as mentioned above.

The second half of the hypothesis, that boys are more susceptible than girls to psychosocial stress, necessitates a broader consideration of sex-linked susceptibilities. Although boys are stronger than girls in muscle power (24), they more readily capitulate under biological stress. There is sound evidence that boys are more susceptible than girls to almost any kind of physical hazard: for example, complications of pregnancy and childbirth (9, 61), infections (67), and radiation effects (11, 26). Except for auto-immune diseases, most illnesses in childhood where there is a sex difference are commoner in boys and mortality rates are also higher (5, 10). Why this should be is not known, but it is likely that the sex difference in maturation is one factor. Girls are more mature than boys at all ages up to maturity. The difference is as much as several weeks even at birth, is one year at the time of starting school and almost two years at puberty (20, 63). This difference is due to the presence of the *Y* chromosome (64). The sex difference is even more marked with *disorders* of development (such as delays in language) than it is with normal development (54).

Whatever the reason, however, there can be no doubt that in terms of vulnerability to physical stresses the male is the weaker sex. I suggest that he may also be more vulnerable to psychological and social stresses. The issue has been astonishingly little investigated. Apart from deaths following bereavement where, as I have already mentioned, males seem to be more susceptible, there is very little evidence one way or the other. People have suggested previously that boys are more liable to the ill-effects of family discord (68), but the opposite has also been argued (2, 65). Unfortunately most of the studies which might have provided crucial information on the issue did not examine sex differences. For example, Yarrow (72), in his review of the effects on children of separation experiences, noted that scarcely anyone had examined sex differences. I hope that sex differences may receive more attention in the future.

This train of thought has taken us rather a long way from Dr Robins's review of follow-up studies, although it was her finding on sex differences and the lack of mention of this in her review which prompted these comments. That I had to focus on a few highly specific areas of dispute is a reflection of my very large measure of agreement with what she had to say.

REFERENCES

1. Barker, P., and Fraser, I. A. (1968). 'A controlled trial of halperidol in children', *Brit. J. Psychiat.* **114**, 855–7.
2. Baruch, D. W., and Wilcox, J. A. (1944). 'A study of sex differences in preschool children's adjustment coexistent with interparental tensions', *J. genet. Psychol.* **64**, 281–303.
3. Becker, W. C., Madsen, C. H., Arnoll, C. R., and Thomas, D. R. (1967). 'The contingent use of teacher attention and praise in reducing classroom behaviour problems', *J. spec. Educ.* **1**, 287–97.
4. Bell, R. Q. (1968). 'A reinterpretation of the direction of effects in studies of socialization', *Psychol. Rev.* **75**, 81–95.
5. Bentzen, F. (1963). 'Sex ratios in learning and behaviour disorders', *Amer. J. Orthopsychiat.* **33**, 92–8.
6. Bronson, W. C. (1966). 'Early antecedents of emotional expressiveness and reactivity control', *Child Develop.* **37**, 793–810.

7. —— (1967). 'Adult derivatives of emotional expressiveness and reactivity control: developmental continuities from childhood to adulthood', ibid. **38**, 801–17.

8. Brown, G. W., and Rutter, M. (1966). 'The measurement of family activities and relationships: a methodological study', *Human Relations*, **19**, 241–63.

9. Butler, N. R., and Bonham, D. G. (1963). *Perinatal Mortality* (Edinburgh and London: E. and S. Livingstone).

10. Childs, B. (1965). 'Genetic origin of some sex differences among human beings', *Pediat.* **35**, 798–812.

11. Conard, R. A., Huggins, C. E., Cannon, B., Lowrey, A., and Richards, J. B. (1957). 'Medical survey of Marshallese two years after exposure to fall-out radiation', *J. Amer. med. Ass.* **164**, 1192–7.

12. Conners, C. K., Eisenberg, L., and Barcai, A. (1967). 'Effect of dextroamphetamine on children: studies on subjects with learning disabilities and school behaviour problems', *Arch. gen. Psychiat.* **17**, 478–85.

13. Craig, M. M., and Glick, S. J. (1965). *A Manual of Procedures for Application of the Glueck Prediction Table* (London: University of London Press).

14. Cunningham, M. A., Pillai, V., and Rogers, W. J. B. (1968). 'Haloperidol in the treatment of children with behaviour disorders', *Brit. J. Psychiat.* **114**, 845–54.

15. Douglas, J. W. B., Ross, J. M., and Simpson, H. R. (1968). *All Our Future: A Longitudinal Study of Secondary Education* (London: Peter Davies).

16. Eisenberg, L. (1968). 'Psychopharmacology in Childhood: a Critique', in *Foundations of Child Psychiatry*, ed. Miller, E. (London: Pergamon Press).

17. —— (1968). 'The interaction of biological and experiential factors in schizophrenia', in *The Transmission of Schizophrenia*, eds. Rosenthal, D., and Kety, S. S. (London: Pergamon Press).

18. —— Conners, K., and Sharpe, L. (1965). 'A controlled study of the differential applications of out-patient psychiatric treatment for children', *Jap. J. Child Psychiat.* **6**, 125–32.

19. —— Lachman, R., Molling, P. A., Lockner, A., Mizelle, J. D., and Conners, K. (1963). 'A psychopharmacologic experiment in a training school for delinquent boys', *Amer. J. Orthopsychiat.* **33**, 431–47.

20. Flory, C. D. (1935). 'Sex differences in skeletal development', *Child Develop.* **6**, 205–12.

21. Forssman, H., and Thuwe, I. (1966). 'One hundred and twenty children born after application for therapeutic abortion refused', *Acta psychiat. scand.* **42**, 71–88.

22. Gardner, G. G. (1967). 'The relationship between childhood neurotic symptomatology and later schizophrenia in males and females', *J. nerv. ment. Dis.* **144**, 97–100.

23. —— (1967). 'Role of maternal psychopathology in male and female schizophrenics', *J. cons. Psychol.* **31**, 411–13.

24. Goss, A. M. (1968). 'Estimated versus actual physical strength in three ethnic groups', *Child Develop.* **39**, 283–90.

25. Gelfand, D. M., and Hartman, D. P. (1968). 'Behaviour therapy with children: a review and evaluation of research methodology', *Psychol. Bull.* **3**, 204–15.

26. Greulich, W. W., Crismon, C. S., and Turner, M. L. (1953). 'The physical growth and development of children who survived the atomic bombing of Hiroshima or Nagasaki', *J. Pediat.* **43**, 121–45.

27. Hoznik, M. P. (1967). 'Environmental correlates of mental growth: prediction from the family setting at twenty-one months', *Child Develop.* **38**, 337–64.

28. Jones, H. G. (1968). 'Behaviour Therapy and Conditioning Techniques', in *Foundations of Child Psychiatry*, ed. Miller, E. (London: Pergamon).

29. Kagan, J., and Moss, H. A. (1962). *Birth to Maturity: A Study in Psychological Development* (New York: J. Wiley).

30. Kety, S. S., Rosenthal, D., Wender, P. H., and Schulsinger, F. (1968). 'The types and prevalence of mental illness in the biological and adoptive families of adopted schizophrenics', in Rosenthal and Kety, *The Transmission of Schizophrenia.*

31. Kringlen, E. (1968). 'The epidemiological-clinical twin study on schizophrenia', ibid.

32. Lebovits, B. Z., Shekelle, R. B., Ostfeld, A. M., and Oglesby, P. (1967). 'Prospective and retrospective psychological studies of coronary heart disease', *Psychosom. Med.* **29**, 265–72.

33. Leff, R. (1968). 'Behaviour modifications and the psychoses of childhood', *Psychol. Bull.* **3**, 396–409.

34. Lovibond, S. H. (1964). *Conditioning and Enuresis* (London: Pergamon Press).

35. Marks, I. M., and Gelder, M. G. (1966). 'Different ages of onset in varieties of phobia', *Amer. J. Psychiat.* **123**, 218–21.

36. McCord, W., and McCord, J. (1959). *Origins of Crime* (New York: Columbia University Press).

37. MacMahon, B. (1968). 'Gene-environment interaction in human disease', in Rosenthal and Kety, *The Transmission of Schizophrenia.*

38. Meduick, S. A., and Schulsinger, F. (1968). 'Some premorbid characteristics related to breakdown in children with schizophrenic mothers', ibid.

39. Molling, P. A., Lockner, A. W., Sauls, R. J., and Eisenberg, L. (1962). 'The impact of perphenazine and placebo on committed delinquent boys', *Arch. gen. Psychiat.* **7**, 70–6.

40. Moore, T. (1967). 'Language and intelligence: a longitudinal study of the first eight years. Part I: patterns of development in boys and girls', *Hum. Develop.* **10**, 88–106.

41. —— (1968). 'Language and intelligence: a longitudinal study of the first eight years. Part II: environmental correlates of mental growth', ibid. **11**, 1–24.

42. Nuffield, E. J. A. (1961). 'Neurophysiology and behaviour disorders in epileptic children', *J. ment. Sci.* **107**, 438–58.

43. Power, M.J., and Shoenberg E.(1967). *Children before the Courts in one Community* (Mimeographed Report).

44. Pritchard, M., and Graham, P. (1966). 'An investigation of a group o patients who have attended both the child and adult departments of the same psychiatric hospital', *Brit. J. Psychiat.* **112**, 603–12.

45. Rees, W. D., and Lutkins, S. G. (1967). 'Mortality of bereavement', *Brit. med. J.* **4**, 13–16.

46. Robins, L. N. (1966). *Deviant Children Grown Up* (Baltimore: Williams and Wilkins).

47. Rosenthal, D. (1962). 'Familial concordance by sex with respect to schizophrenia', *Psychol. Bull.* **59**, 401–21.

48. —— Wender, P. H., Kety, S. S., Scuhlsinger, F., Welner, J., and Østergaard, L. (1968). 'Schizophrenics' offspring reared in adoptive homes', in Rosenthal and Kety, *The Transmission of Schizophrenia.*

49. Rutter, M. (1966). *Children of Sick Parents* (London: Oxford University Press).

50. —— (1967). 'A children's behaviour questionnaire for completion by teachers: preliminary findings', *J. Child Psychol. Psychiat.* **8**, 1–12.

51. —— (1970). 'Sex differences in children's responses to family stress' (in press).

52. —— and Brown, G. W. (1966). 'The reliability and validity of measures of family life and relationships in families containing a psychiatric patient', *Social Psychiatry,* **1**, 38–52.

53. —— and Graham, P. (1966). 'Psychiatric disorder in 10 and 11 year old children', *Proc. roy. Soc. Med.* **59**, 382–7.

54. —— Graham, P., and Yule, W. (1970). *Neurological Disorders in Childhood: A Study of a Small Community* (London: Wm. Heinemann).

55. —— Greenfeld, D., and Lockyer, L. (1967). 'A five to fifteen year follow-up study of infantile psychosis. II: social and behavioural outcome', *Brit. J. Psychiat.* **113**, 1183–99.

56. —— Tizard, J., and Whitmore, K. (Eds.) (1970). *Education, Health and Behaviour* (London: Longmans Green).

57. —— Yule, W., Tizard, J., and Graham, P. (1967). 'Severe reading retardation: its relationship to maladjustment, epilepsy and neurological disorders', *Proc. Int. Conf. Assoc. Spec. Educ. London, July 1966* (Stanmore, Middx.: Assoc. Spec. Educ.).

58. Shaffer, D., Costello, A. J., and Hill, I. D. (1968). 'Control of enuresis with imipramine', *Arch. Dis. Childh.* **43**, 665–71.

59. Shepherd, M., Oppenheim, A. N., and Mitchell, S. (1966). 'Childhood behaviour disorders and the child guidance clinic: an epidemiological study', *J. Child Psychol. Psychiat.* **7**, 39–52.

60. Shields, J. (1968). 'Summary of the genetic evidence', in Rosenthal and Kety, *The Transmission of Schizophrenia.*

61. Singer, J. E., Westphal, M., and Niswander, K. R. (1968). 'Sex differences in the incidence of neonatal abnormalities and abnormal performance in early childhood', *Child Develop.* **39**, 103–12.

62. Sprague, R. L., and Werry, J. S. (eds.) (1968). '*Survey of Research on Psychopharmacology of Children* (Mimeographed Report, Children's Research Center, University of Illinois).

63. Tanner, J. M. (ed.) (1960). *Human Growth* (London: Pergamon Press).

64. —— Prader, A., Habich, H., and Ferguson-Smith, M. A. (1959). 'Genes on the *Y* chromosome influencing rate of development in man: skeletal age studies in children with Klinefelter's (*XXY*) and Turner's (*XO*) syndromes', *Lancet,* **2**, 141–4.

65. Toby, J. (1957). 'The differential impact of family disorganisation', *Amer. Sociol. Rev.* **22**, 505–12.

66. WARING, M., and RICKS, D. (1965). 'Family patterns of children who become adult schizophrenics', *J. nerv. ment. Dis.* **140**, 351–64.

67. WASHBURN, T. C., MEDEARIS, D. N., and CHILDS, B. (1965). 'Sex differences in susceptibility to infections', *Pediat.* **35**, 57–64.

68. WEINSTEIN, E. A., and GEISEL, P. N. (1960). 'An analysis of sex differences in adjustment', *Child Develop.* **31**, 721–8.

69. WEISS, G., WERRY, J., MINDE, K., DOUGLAS, V., and SYKES, D. (1968). 'Studies on the hyperactive child: V. The effects of dextro-amphetamine on behaviour and intellectual functioning', *J. Child Physiol. Psychiat.* **9**, 145–56.

70. WENDER, P. H., ROSENTHAL, D., and KETY, S. S. (1968). 'A psychiatric assessment of the adoptive parents of schizophrenics', in Rosenthal and Kety, *The Transmission of Schizophrenia.*

71. WING, L., BRAMLEY, C., HAILEY, A., and WING, J. K. (1968). 'Camberwell cumulative psychiatric register. Part I: Aims and methods', *Social Psychiatry*, **3**, 116–22.

72. YARROW, L. (1964). 'Separation from parents during early childhood', eds. Hoffman, L. W., and Hoffman, M. L. (New York: Russell Sage Foundation).

2

J. W. B. DOUGLAS

MRC Unit
London School of Economics, London

Dr Robins has admirably described the contributions that follow-up studies can make to our understanding of behaviour disorders in children. She has, however, emphasized the positive contributions and has hardly touched on the practical problems that beset any longitudinal study. The following comments are stimulated by my own experience of a cohort study, the 1946 National Survey of some 5,000 births.

In any longitudinal inquiry it is essential to control for the possibility that the sample may be altered through the fact that it is the subject of study. Medical examinations, discussions with mothers, doctors, and teachers, psychological tests, and other research procedures may result in a sophistication and self-consciousness of the sample members which, as time passes, make them increasingly less representative of a normal population. The more intense the study the greater the danger of altering attitudes.

This danger has been widely recognized but I know of no other inquiry than the National Survey which has attempted to control for it. In this study I kept a group of 1,200 families in 'cold storage' and gathered information about them from records of schools, court appearances, and employment. One direct contact with them was made when they were 11 years old, when they were given a medical examination and their homes were visited. So far only very minor distortions have been found: while all serious speech and eye defects were treated in both the controls and the main survey sample, minor defects were more likely to be treated for the main survey children. In all other respects we have been unable to detect any distortion, though we suspect we may find in the long run that we have influenced the employment of some of the members. The early school-leavers were seen by youth employment officers on our behalf and some of these in their reports said that they felt the young person they had seen was unsuitably placed and that they were doing what they could to get him more satisfactory employment.

These results are encouraging, but the National Survey is a study in which contacts were relatively infrequent (on the average once every eighteen months) and we might well have got a different answer if the group had been more intensively studied. If the study had concentrated primarily on behaviour disorders in children it would have been very difficult to avoid altering the mothers' perception of their children and their behaviour.

Losses are, of course, one of the major troubles of follow-up studies. With local studies losses can be very high and, apart from the waste of effort they represent, can introduce serious biases. Those who leave are always different from those who stay, and this holds both for families who move out of the area of study and for families who refuse. In the Newcastle 'Thousand Family Study', 12 per cent of families were lost during the first year of the survey, a high proportion being in the Registrar-General's classes I and II. Losses can be greatly cut down if the study covers either a complete country or a regional area chosen to minimize migration across its boundaries. In the National Survey, for example, we have been able to keep losses at approximately 10 per cent of those living in the country and we know the bare bones of employment history and educational achievement for 98 per cent of the sample. On the surface the losses from the National Survey look relatively unbiased; the actual social structure of the sample, for example, does not differ significantly from what would have been expected if loss had been random. But there is some evidence that the families we have lost include a high proportion of the unhappily married and of those that have children who are backward or in some way not doing their parents credit.

In a longitudinal study it is always difficult to decide how frequently to visit or test. Each contact carries with it the possibility of refusal and it is tempting to cut them for this reason to a minimum and try to extract the maximum information on each occasion. This seems to me a mistaken policy, particularly when using untrained interviewers. We have managed to maintain contact with the National Survey members at a minimum of two-yearly intervals and at some points more frequently than this. The only time when we felt we were making too many contacts was in the second primary school year, when the homes were visited three times. Both the mothers and the health visitors, who were acting as our agents, became surfeited and losses began to rise. One of the most important aspects of a longitudinal study, as Dr Robins has pointed out, is that it enables us to date events and so escape from the retrospective interpretations so often made when children have behaviour troubles which their parents try to explain in the context of popular theory. Eighteen-month intervals are rather wide when it comes to recording many events and looking back I would have liked to have been able to visit the homes every six months; I think this would have been feasible only if we had had a small and really docile population.

One of the limitations of follow-up studies is that their findings apply with certainty only to the group which is being followed up. For example, when we look at the relationship between early care and present behaviour, we are concerned with types of upbringing that were in vogue twenty years ago. These will certainly differ from present methods of care in many ways, some of which may be subtle and difficult to identify, and it is doubtful how far we can use the associations found in one generation to predict those that are likely to be found in the next.

Longitudinal studies such as the National Survey, which are large and use relatively unskilled interviewers, have been criticized as giving no adequate assessments of behaviour and personal characteristics. It is clear that sophisticated tests cannot be used in this type of study but I think these criticisms are misplaced because they ignore the number of different sources of information on which it is possible to draw. Assessments can be made by mothers, teachers, and school doctors; the children can fill in inventories; events such as truancy, delinquent acts, frequent changes of job, illegitimate babies, and marital problems reflect the stability of the survey members. Lastly there are admissions to hospitals for psychiatric care or attendance at clinics. All these taken individually are open to criticism but together they provide evidence of the subject's ability to deal with everyday problems of life, marriage, and employment, and may give a more realistic assessment of adjustment than the tests of personality that we might have given if we had had a smaller and more manageable sample.

Perhaps the most useful aspect of longitudinal studies based on national populations is that they offer the opportunity to look at the ways in which successive generations of parents tackle the problems of bringing up children. We are at the moment planning a second generation study with the aim of seeing how experiences in the early life of the survey members are related to the way in which they themselves bring up their children.

A series of short cohort studies covering the early years of childhood would show how methods of child care are influenced by current theories. Unfortunately this is just the sort of information that is not available from the national studies so far available. The 1946 Survey, though it collected some information about the early years, was not planned specifically as a study of behaviour. In the early years it was directed to problems of hospitalization and childhood illness. The succeeding 1958 cohort study has a great gap covering the whole of the first seven years.

I have outlined some of the difficulties of planning longitudinal studies. In doing so I do not want to diminish their value. It seems to me that the most important element in making them feasible in the future would be to have a good system of record linkage. We could then study intensively behaviour and experience at certain ages and wait to take up the story later on.

3

STANDARDIZATION OF CLINICAL ASSESSMENT

A standard form of psychiatric present state examination
and a method for standardizing the classification of symptoms

J. K. WING
MRC Social Psychiatry Research Unit,
Institute of Psychiatry, University of London

This paper describes the principles underlying the construction and use of a technique for case-identification and description which involves the partial standardization of the 'Present State Examination' (PSE) undertaken by psychiatrists when they first see a patient. Certain rules are then applied in order to classify the symptoms and signs so elicited. These procedures are in use in several parts of the world and some results will be described later by Dr J. E. Cooper and Dr Norman Sartorius. My task is to describe the reasons for developing the PSE, which were partly technical and partly theoretical. The need for a reliable method of case identification is obvious enough to anyone interested in epidemiology. All the well-known surveys have been based upon interviews. In the classical Scandinavian studies these were solidly clinical, that is, the doctor himself saw the respondents and decided whether a psychiatric disorder was present or not. The interview was ordinarily quite unstandardized but it was not unskilled. The basic checklist of symptoms was carried in the examiner's head and his technique of interviewing was improvised to meet the occasion. The principles of classification were also rough and ready. The whole process sounds quite lamentably deficient in consistency, reproducibility, and comprehensibility, but it did get results. These studies were based on the so-called 'medical model', that is, the investigators thought they were counting the frequency at which psychiatric illness occurred in the population.

More recent investigators have not started with the same confidence in an underlying disease theory. The checklists used in the Midtown, Manhattan, and Stirling County surveys were not constructed with the idea of detecting specific illnesses. It is sometimes difficult to know what the items were designed to elicit. Take, for example, the question—'Are you ever troubled by your hands feeling damp: often, sometimes?' 'Are you' refers to the present. 'Ever' can be construed as referring to the past, or to recent events, or to the present moment, and implies that the symptom must be either present or absent. The term 'troubled' leaves all interpretation to the informant. 'Often, sometimes' is a crude measure of frequency, which gives a confusing effect following the word 'ever'. The informant can respond to any one component of this question, or to a mixture of several, and thus a large element of ambiguity and imprecision is allowed. The interviewers were not usually skilled, in the sense that they could undertake a cross-examination and then make informed judgements as to whether symptoms were present or not, in the way the Scandinavian investigators could, and the subsequent clinical evaluation of the checklists could do very little to retrieve the situation. Although the terminology of psychiatric nosology is used to some extent, it is clear that the investigators were really interested in one underlying and all-embracing dimension, with 'Health' at one pole and 'Pathology' at the other. Thus the item about damp hands forms part of an over-all evaluation of deviation from health, not part of a specific examination to discover whether the individual could be diagnosed as suffering from an anxiety state.

This dimensional approach has been more explicitly expounded and more thoroughly practised by psychologists, who also used checklists of items to measure clinical condition (3). Again, the method of collecting information is largely uncontrolled and can be unskilled, but a great deal of precision and sophistication is introduced into the statistical treatment of the ratings. The technique can be very reliable. Diseases are replaced by factors or clusters such as 'Hostile Paranoid' or 'Anxious-Disorganized'. There is no particular reason why such constructs should have a specific aetiology or course or outcome, or indicate a need for a specific treatment, and epidemiologists have understandably not used them. However, in some parts of the world a diagnosis is not thought to carry these implications either, and in fact is not taken seriously at all. In such places the use of instruments such as the Lorr scale would appear to be superior to clinical procedures, at least for purposes of medical taxonomy.

Two attempts have been made to bring the interview itself under control, in addition to standardizing the checklist. Spitzer and his colleagues developed an instrument known as the Mental Status Schedule (MSS) which has been extensively used and tested (5, 6). Analysis has followed both dimensional and diagnostic lines (7, 8). Our own approach has been technically very similar: that is, we use a standard checklist, standard definitions, a structured interview, and a skilled interviewer (10). The main difference lies in the thinking which determines the selection of items and the conduct of the interview, itself reflecting some fairly characteristic differences between British and American clinicians. We ask more specific questions and many more of them, on the assumption that they might be important in differential diagnosis. For example, it is said that voices talking *about* a patient do not have the same diagnostic significance as voices speaking *to* him. I do not know whether this is true but it cannot be tested unless the questions are asked. In addition, we use 'cross-examination' very much more, not being content to accept a patient's 'yes' or 'no' to a standard question as the only criterion for rating. Fortunately, the U.S./U.K. Diagnostic Project was able to use both the MSS and the PSE in its studies and so it will be possible to discover the advantages and disadvantages of each.

Before considering the principles on which the PSE is based, it might be wise to digress a little on the subject of the 'medical model'.

The concept of illness

The diagnosis of disease is the basis of modern medicine. Without it, the scientific progress which has occurred within the past hundred years, and particularly during this century, could not have occurred. Medicine would still be in the dark ages. It is not surprising that psychiatrists have endeavoured to recognize and classify illnesses and that this attempt should still be regarded as one of their most important scientific activities.

Classically, an illness has one or a few necessary causes which lead, under certain circumstances, to biological dysfunctions (defined in terms of some theory of normal functioning). These dysfunctions in turn give rise to a characteristic pathology and to a recognizable pattern of symptoms and signs with a characteristic course and outcome. Knowledge of the cause allows one to attempt prevention or treatment. Knowledge of the course gives clues about management. At the very least, it is possible to give a prognosis. There may, of course, be numerous other factors—biological, psychological, or social—which influence resistance and susceptibility, precipitate or delay onset or relapse, produce or prevent complications, modify or obscure symptoms, and so on.

A diagnosis is a hypothesis that a disease is present and that other diseases are absent. It can be based on only partial knowledge. In psychiatry, it has to be, since very little is known of the causes, dysfunctions, or pathology of most conditions. However, the hypothesis can always be tested, even if only crudely, since a diagnosis should still imply a course, a treatment, and a means of management. There is little doubt that the concept of illness can be applied successfully to the organic psychoses. So far it has been much less successful with the personality disorders. The following remarks apply only to the 'functional' psychoses and neuroses. Those who criticize the 'medical model' as applied to these conditions do so on the grounds that no adequate studies have been carried out to show that 'schizophrenia', 'anxiety state', and other diagnostic constructs, actually conform to the classical picture. I do not know of a substantial critique which has actually attempted to survey the literature in any detail, and many authors try to take their argument very much further than it will go. It is commonly forgotten that the concept of disease must be supplemented by the concept of handicap (which is also part of the medical model), just as rehabilitation and management must supplement treatment.

Purely dimensional theories appear to some to offer an alternative, not only for the personality disorders, but for the neuroses and psychoses as well. As applied to the latter they are, at the moment, pale imitations of disease theories. Statistical constructs, such as 'Perceptual Disorganization', have to be treated as diseases, or at least as symptom clusters, if they are to be useful at all (except perhaps to measure change), and no one has claimed that it is possible to use them in this way. Conceivably, it may one day be possible to describe a large number of normally distributed dimensions in such a way that the calculation of a patient's unique position in multi-dimensional space will give the same information about treatment, course, and outcome that correct diagnosis would. There is very little evidence as yet in favour of this view. Meanwhile, we must sharpen the diagnostic tools we have, try to achieve reliability and consistency when using them, and thus prepare the way for tests of the validity of the diagnostic system. We should not assume that diagnoses are useful *only* because generations of psychiatrists have thought so, nor should we assume that this is the only fruitful model to be applied by psychiatrists.

Considerations of this kind led to the construction of the PSE and to an attempt to provide a standard classification based on PSE symptoms. Making these ideas explicit may lead some to feel that so many preconceptions are built-in to the instrument and subsequent classifying procedures that anyone who uses them is bound to be influenced. This is true, of course, but it is equally true of all instruments in the field that are worth using.

The principles of the interview

The basis of the interview is a checklist of some 600 items which systematically covers all the phenomena likely to be considered during a clinical examination of present state. Each item is defined in greater or lesser detail and ratings are coded in numerical form. Most items are worded as questions so that it would be possible to conduct the interview without departing from the schedule at all, though I doubt whether this ever occurs. The items are grouped together in sections. Three main principles guide the conduct of the interview:

(*a*) The interview is based on, and retains the main features of, an ordinary clinical examination. Each item represents a trait and the definitions, examples, training video-tapes, and instruction manuals are intended to provide the examiner with an idea of what that trait is supposed to be, so that he can make a judgement as to whether it is present or not. Many of the questions laid down are followed by extra probes but the examiner is free to ask as many questions as he likes, in order to establish whether the item should be rated positively. The wording of most of the questions is based on the way patients have actually described their symptoms, but further questions are often required, particularly when a patient responds very briefly. He is asked to describe the experience in his own words and his responses may then need further clarification. This is the process of cross-examination, a basic part of the clinician's skill, and well accepted by patients. It means that the wording and the order of the questions in the schedule are not immutable. The schedule is nothing at all like a questionary and the whole procedure is extremely flexible. The essence of the first principle is that the *clinician* decides whether the symptom is present and, if so, to what degree.

(*b*) A second clinical feature is that most sections have 'cut-off points', enabling the interviewer to ask a few questions about each major area of psychopathology, but to conduct a detailed examination only in those which seem at all likely to be positive.

(*c*) Thirdly, the examination is based upon the patient's account of his experiences during the previous month. This is a very important limitation because it excludes most of the psychiatric history (which is the subject of a separate inquiry). Clearly, diagnoses which depend upon historical information, whether previous episodes (such as 'manic-depressive psychosis') or aetiology (such as *delirium tremens*) or response to social situations over a long period of time ('personality disorders'), cannot be based solely on present-state information.

Scoring

There are several alternative methods of scoring, depending on how ratings of items are used and combined. Several large-scale analyses are being carried out at the moment which will help to decide what the most useful patterns are. Profiles of 'syndrome scores' (see section below on diagnosis) are given in the Appendix, separately for patients with mania (296.1), with psychotic depression (296.2), and with paranoid schizophrenia (295.3). Many other profiles could be given—by length of history, age, sex, family relationships, outcome, and so on.

Editions

The PSE schedule has passed through several editions, though its basic principles have not changed. The main ones being analysed at the moment are the seventh and eighth editions (used by the MRC and by the U.S./U.K. Project) and the WHO edition which is a shorter version of the eighth focused mainly on psychotic symptoms. There is also one major project just completed which used the sixth edition. A ninth edition, and a brief version, will be based on the results of the current analysis.

Reliability

In spite of the flexibility with which the instrument can be used, the results obtained by trained observers are reasonably reliable (1, 10). This is true of ratings made simultaneously during the interview by the examiner and an observer, and of ratings of audio- or video-tapes, and of consecutive interviews undertaken by different examiners within a few days of each other.

Standardization of the processes of diagnosis

A diagnosis of one of the functional psychoses or neuroses is usually made on the basis of three kinds of information:

(*a*) Symptoms present during the current episode of illness.

(*b*) Symptoms present during past episodes, and the general course of the illness.

(*c*) Biological, psychological, or social factors thought to have caused or precipitated one or more episodes.

The first kind of information is most easy to acquire, the second and third are increasingly difficult. The particular mixture of the three which any one clinician uses in a given case is unique. This is why ordinary diagnosis is so unreliable in statistical terms.

The eighth revision of the *ICD* and the various national glossaries that go with it, act to some extent towards standardization of practice but there are still no really operational definitions or unambiguous principles of classification (11).

For research purposes, we can afford to separate the three kinds of information from each other, classifying firstly on the basis of symptoms elicited in the PSE, then adding information concerning past episodes, and finally adding judgements about aetiology. The following account of the first stage of this process is based on work being carried out jointly by myself and Dr John Cooper. A computer program (named Catego for short) has been written and tested by Mrs Cynthia Taylor (MRC Computer Services Centre).

Classifying PSE ratings

In the first stage, PSE items are grouped, on the basis of clinical judgement, into 'symptoms', of which there are 145. Thus symptom number 52, 'poor concentration', is composed of four items from the eighth edition of the PSE:

9.B1 Do your thoughts drift off when you try to concentrate?

9.B2 Can you concentrate on a television (radio, film) programme all the way through?

9.B3 Can you concentrate on something you read, such as something in the paper? Can you read it right through?

9.B4 Can you concentrate on what people are saying—take it in; or do you lose the thread of the conversation?

Since each of these items can be rated 0, 1, or 2 the total possible score for the symptom is 8, and all symptoms have such a 'symptom score'. For the purposes of classification, a threshold point is allocated to each of the 145 possible symptoms which defines whether it is to be regarded as present or absent. In this case the threshold is 3 and symptom number 52 is only regarded as present if the score is 3 or more.

In the second stage of the Catego program, those symptoms which are defined as present are combined into 'syndromes', of which there are 35. Again, decisions as to how symptoms should be combined are made clinically. For example, syndrome number 10, 'Hypomania' is composed of the following seven symptoms:

Symptom no. 44. Subjective euphoria.
„ 45. Ideomotor pressure.
„ 46. Grandiose ideas.
„ 74. Too much speech.
„ 75. Distractibility.
„ 84. Hypomanic affect.
„ 137. Hypomanic content of speech.

There is, of course, a 'syndrome score' consisting of the sum of symptom scores, but, for the purposes of Catego, a 'syndrome certainty' is allocated, on the basis of the number of symptoms present. In this example, one symptom is called 'hypomania', two symptoms are called 'hypomania +', and three symptoms are called 'hypomania ++'. All four versions of the PSE reach this common point and the rest of Catego is based upon the 35 syndromes and their degree of certainty.

The process of combining syndromes in order to produce a small number of 'provisional diagnostic categories' is a complicated one, which will be described in detail elsewhere. Certain decisions are relatively easy. If syndrome number 1 ('nuclear schizophrenia') is present with the highest degree of certainty (++), which means that symptoms such as 'voices discussing patient in third person' and 'delusions of control' are present, one of the categories allocated must be a schizophrenic one. Similarly, if the 'catatonic syndrome' or the 'obsessional syndrome' or the 'situational anxiety syndrome' are definitely present, a category is indicated—whatever other syndromes might also be there.

The difficulty comes with syndromes such as 'religious and grandiose delusions' or 'overactivity', which can suggest several different diagnostic categories. Everything then depends upon the context of other syndromes. If these two are present, together with 'hypomania', and if other syndromes are not present, allocation to a category ('mania') is easy. If other syndromes (for example, 'slowness' or 'auditory hallucinations') are also present, a single category is unlikely to do justice to the clinical picture. Clinical judgements are built in to all stages of the program and in stage 10 a set of categories is printed out, together with the degree of certainty allocated to each. A patient could conceivably be given a category indicating schizophrenia,

TABLE 3.1. ICD (*eighth revision*) *classification by clinicians, and by the Catego program*

		Seventh edition and Clinicians A and B			*Eighth edition Clinician C*		
		No.		%	*No.*		%
Same category							
Catatonic schizophrenia	(295·2)	1	125	56·1	—	92	76·7
Paranoid schizophrenia	(295·3)	102			70		
Residual schizophrenia	(295·6)	2			1		
Schizo-affective psychosis	(295·7)	7			—		
Mania	(296.1)	2			5		
Psychotic depression	(296.2)	5			2		
Non-psychotic depression	(300.4)	6			14		
Similar category							
295·3 and 295·7		19	40	17·9	8	14	11·7
Other schizophrenic groups		21			8		
Paranoid state (297·9)							
297·9 or 299 with 295·0–9		20	21	9·4	4	4	3·3
297·9 with 296.1		—			—		
297·9 with 296.2		1			—		
Psychotic and non-psychotic conditions							
295·0–9 and non-psychotic		18	24	10·8	3	3	2·5
299 and non-psychotic		2			—		
297·9 and non-psychotic		4			—		
Somewhat similar category							
295·7 and 296.1		3	6	2·7	1	3	2·5
295·7 and 296.2		3			2		
Different categories							
295·0–9 and 296.1		4	7	3·1	1	4	3·3
295·0–9 and 296.2		1			1		
296.2 and 300.4		2			2		
Total		223		100·0	120		100·0

1. These two projects were primarily concerned with schizophrenia but other types of patient were examined during the course of screening. J. L. T. Birley and J. K. Wing were the examiners using the seventh edition and J. Leff used the eighth.

2. The table does not show which categories were allocated by the clinician and which by Catego.

3. The Catego program deals only with present symptomatology and is not, therefore, intended to result in a diagnosis. It gives more information when one final category is not required. (For example, one patient was allocated to three categories: 'agitated psychotic depression ++', 'delusional schizophrenia?', and 'obsessional neurosis +', but for the present comparison was classified as 'psychotic depression' only. The clinician's diagnosis was schizo-affective psychosis.)

4. The table shows the extent to which a crude classification based on the Catego procedure is similar to the diagnosis given by clinicians and therefore indicates to what extent the latter are using data other than present symptomatology.

In the project using the seventh edition, 17 patients with a clinical diagnosis of paranoid state and 4 with unspecified psychosis were included. These diagnoses were made because it was not possible to come to a definite conclusion about the role of factors such as alcohol, amphetamine, epilepsy, or minimal brain damage. In another 24 cases the patient had no psychotic symptoms at the time of examination but the clinician diagnosed a psychotic condition on the basis of the history. Thus the differences between Catego and the clinician are due to the inclusion of historical information by the latter in at least 45 cases out of 223.

In the project using the eighth edition, cases with possibly complicating factors such as alcoholism were excluded. The patients all had psychoses of recent onset and there were few without psychotic symptoms. At least 7 out of 120 were classified differently by Catego and the clinician because of the use of historical data by the latter, but probably not many more.

5. The high over-all agreement is mainly due to schizophrenia, which can clearly be diagnosed by these clinicians mainly on the present state; but only about half the patients with mania and psychotic depression were classified in the same way both by the clinician and by the Catego program. This is partly due to historical information being used by the clinician and partly to the use of rules of diagnosis which are not precisely specifiable and therefore difficult to apply without variation from case to case. The advantage of the Catego program is that, once the rules are specified, there will be no variability in application whatsoever.

one indicating mania, one indicating depression of some kind, one indicating an anxiety state, and one indicating obsessional neurosis. In practice most patients are allocated only two or three categories. These are *not* diagnoses; there are still the other two kinds of information, historical and aetiological, to be added before anything like a diagnosis can be made, and we are working on this now.

Probably different diagnostic traditions use different proportions of these three kinds of information. It looks as though the Maudsley tradition is based to quite a large extent on 'present state' data. For example, Dr Birley and I examined a series of 223 patients, using the seventh edition, all of them suffering from a form of psychosis (most from schizophrenia) and all living in a defined area in south-east London. Our diagnoses are compared with the Catego classification (translated into the *ICD* eighth revision diagnostic groupings) in Table 3.1. The fact that 56 per cent are placed into the same groups by Catego and by our clinical judgements, indicates that in this proportion of cases we were not using historical or aetiological data to any considerable extent. In the remainder of the cases, however, we were doing this (except for a proportion where we were classifying present symptomatology inconsistently).

The result of a similar exercise, in which Dr Leff interviewed 120 patients using the eighth edition is also shown: 77 per cent were placed into identical groups by Catego and the clinician. All the discrepancies were explained either by a high weighting being applied to one particular symptom by the clinician, resulting in a different sub-category, or by the use of historical information. (In this exercise patients with alcoholism or mental subnormality and other possibly complicating factors were excluded.)

It would not be expected that other diagnostic traditions would rely so much on present state symptomatology. A preliminary run on 32 PSE interviews by Danish psychiatrists, for example, showed 13 allocated to precisely the same group as Catego and a further 7 to similar categories. Four were placed in somewhat similar categories, and 9 in quite different ones. Consideration of the last group showed that information not in the PSE accounted for the discrepancies. Whether agreement will be increased when historical and aetiological information is included is uncertain. A comparison of my own diagnoses with Catego shows that I tend to use such material rather haphazardly and I suspect that most clinicians do. Practice is therefore likely to vary even more between different centres. The great advantage of the Catego approach is that, once the information has been gathered and specified judgements made, the process of categorization is always precisely the same for each case.

However, the proof of the pudding is in the eating, and the sceptics, particularly perhaps psychologists and sociologists, will want to see evidence of the validity as well as the reliability of such a system of classification.

Numerical taxonomy

Many people who feel that writing a psychiatrist's clinical judgments into a classification programme is simply perpetuating a system of doubtful validity will prefer to use a numerical taxonomy system (4), in which specified mathematical operations are carried out on data such as PSE ratings, in order to produce classes in which members 'are mutually more highly related to one another than they are to non-members'. Clinical judgement is not, of course, excluded, since it is necessary in order to select the variables and make the ratings, but once the numbers have been obtained they are classified solely according to mathematical principles and no further clinical decisions are allowed. If the procedure comes up with a classification which does not conform at least approximately to psychiatric nosology, the process of evaluating validity will be very complicated. The advantage of the system in general clinical use is that there is more than half a century of experience behind it, including quite a substantial scientific literature. If alternatives are to be produced in large numbers, as is quite possible using the new techniques, there will be a danger that most of them will remain untested. It seems reasonable to use clinical diagnosis (particularly if standardized) as a first yardstick.

Conclusion

Although the research worker must remain critical, he must build upon some foundation, however shaky. Popper has taught us how difficult it is to do entirely without hypotheses (or prejudices). Our procedures are imitating ordinary clinical processes throughout. There is nothing in what I have described that is not part of the everyday diagnostic practice of clinical psychiatrists. We have standardized, defined, and made more explicit some of these processes but we have not introduced new ones.

REFERENCES

1. KENDELL, R. E., EVERETT, B., COOPER, J. E., SARTORIUS, N., and DAVID, M. E. (1968). 'The reliability of the Present State Examination', *Social Psychiatry*, **3**, 123.
2. LEIGHTON, D., HARDING, J. S., MACKLIN, D. B., MACMILLAN, A. M., and LEIGHTON, A. H. (1963). *The Character of Danger. III.* The Stirling County Study of Psychiatric Disorder and Sociocultural Environment (New York: Basic Books).
3. LORR, M. (1966). *Explorations in Typing Psychotics* (London, New York: Pergamon Press).

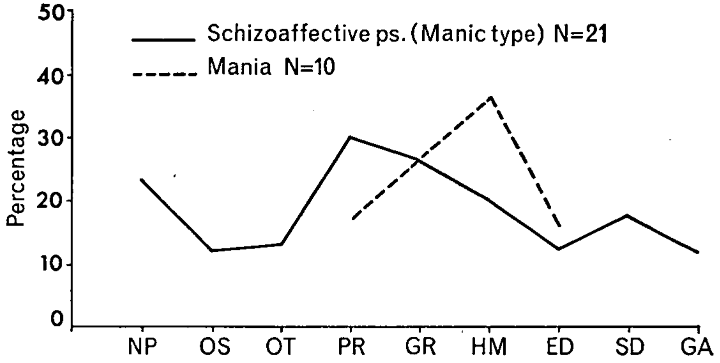
4. Sokal, R. R., and Sneath, P. H. A. (1963). *Principles of Numerical Taxonomy* (San Francisco and London: Freeman).
5. Spitzer, R. L., Fleiss, J. L., Burdock, E., and Hardesty, A. (1964). 'The mental status schedule: rationale, reliability and validity', *Compr. Psychiat.* **5**, 384.
6. —— Endicott, J., and Fleiss, J. L., (1967) 'Instruments and recording forms for evaluating psychiatric status and history: rationale, method of development and description', ibid. **8**, 321.
7. —— Fleiss, J. L., Endicott, J., and Cohen, J. (1967). 'Mental status schedule: properties of factor-analytically derived scales', *Arch. gen. Psychiat.* **16**, 479.
8. —— and Endicott, J. (1969). 'Diagno II: further developments in a computer program for psychiatric diagnosis', *Amer. J. Psychiat.* **125**, 12.
9. Srole, L., Langner, T., Michael, S. T., Opler, M. K., and Rennie, T. A. C. (1962). *Mental Health in the Metropolis* (New York: McGraw-Hill).
10. Wing, J. K., Birley, J. L. T., Cooper, J. E., Graham, P., and Isaacs, A. (1967). 'Reliability of a procedure for measuring and classifying Present Psychiatric State', *Brit. J. Psychiat.* **113**, 499.
11. Wing, L. (1970). 'Comments on the psychiatric section of the *ICD* and the British Glossary', *Psychol. Med.* **1**, in press.

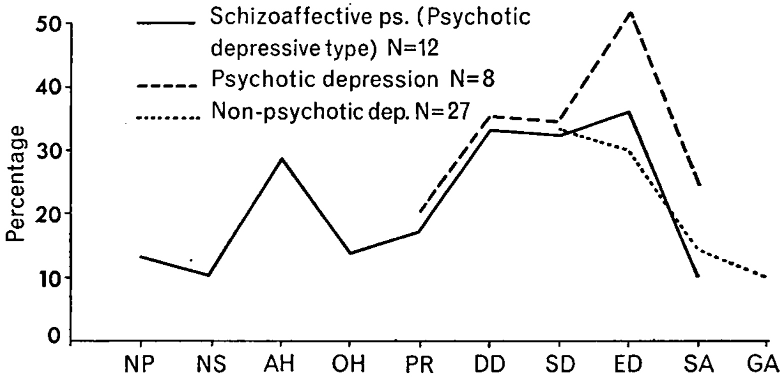
Appendix

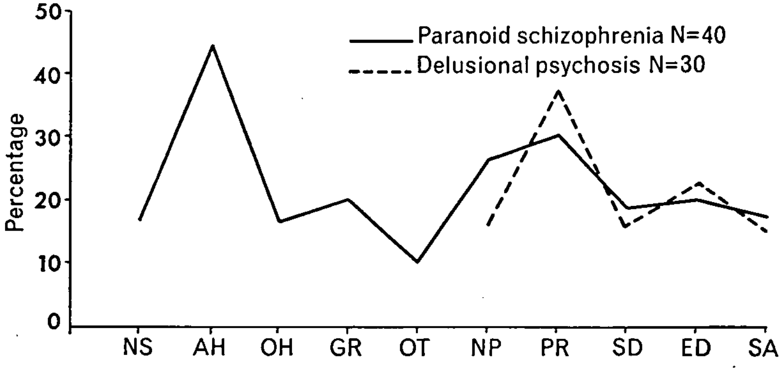
The figure illustrates profiles of syndrome scores for several diagnostic categories derived by the Catego program from PSE ratings. In order to allow comparison of syndrome scores with each other, a 'per cent score' is calculated, the highest score for any patient in the series being made equal to 100. Data from the seventh and eighth editions of the PSE were combined.

Twenty of the 35 syndromes were included in the analysis and are listed below. If the mean per cent score was less than 10, however, that syndrome has been omitted from the figure.

1	NS	Nuclear schizophrenia
2	CS	Catatonic schizophrenia
3	IS	Incoherent schizophrenia
4	RS	Residual schizophrenia
5	DD	Depressive delusions
6	SD	Simple depression
7	ON	Obsessional neurosis
8	GA	General anxiety state
9	SA	Situational anxiety state
10	HM	Hypomania
11	NP	Psychotic mood and perception


Schizoaffective ps. (Manic type) N=21
Mania N=10
Percentage
50
40
30
20
10
0
NP
OS
OT
PR
GR
HM
ED
SD
GA


Schizoaffective ps. (Psychotic depressive type) N=12
Psychotic depression N=8
Non-psychotic dep. N=27
Percentage
50
40
30
20
10
0
NP
NS
AH
OH
PR
DD
SD
ED
SA
GA


Paranoid schizophrenia N=40
Delusional psychosis N=30
Percentage
50
40
30
20
10
0
NS
AH
OH
GR
OT
NP
PR
SD
ED
SA

12	OV	Overactivity
13	SL	Slowness and underactivity
14	SS	Simple schizophrenia
15	AH	Non-specific auditory hallucinations
16	OH	Other non-specific hallucinations
17	PR	Delusions of reference and persecution
18	GR	Grandiose and religious delusions
19	OT	Other delusions
20	ED	Special features of depression

The method of deriving 'diagnostic categories' is explained in the text. When schizophrenic and manic symptoms occur in the same PSE protocol, the category allocated is 'schizo-affective psychosis, manic type'; similarly with a combination of schizophrenic and psychotic depressive symptoms. The category 'paranoid schizophrenia' is allocated when there are auditory hallucinations or 'nuclear' symptoms of schizophrenia. 'Delusional psychosis' is allocated when there are delusions but no nuclear symptoms of schizophrenia or auditory hallucinations and no evidence of 'mania' or 'psychotic depression'. This group is coded with 'paranoid schizophrenia' as 265.3 in the *ICD* eighth revision.

The use of a procedure for standardizing psychiatric diagnosis[1]

J. E. COOPER
Institute of Psychiatry, University of London

This paper describes the use of a set of procedures designed to minimize the diagnostic variation between psychiatrists. An investigation is discussed in which a 'standardized' diagnostic procedure was used to study the sources of variation in mental hospital admission statistics, and particular attention is given to ways of assessing whatever variation remains in the standardized procedures. It is taken as already established that there is a need for such procedures (2, 7, 8, 13).

Before describing the field study, I will discuss the implications of attempting to standardize the psychiatric diagnostic process. The first and most obvious step towards achieving diagnostic agreement between psychiatrists is the production of a set of definitions or a glossary of terms, but any analysis of the components of the diagnostic process immediately makes it clear that the choice of a diagnostic term is only the final stage of a complex process. The process can be divided into four stages and Cooper (2) has described these as follows:

First, the interviewing technique of the psychiatrist; second, the perception of the patient's speech and behaviour; third, the inferences and decisions made by the psychiatrist on the basis of what he has perceived; and fourth, the attachment of a particular diagnostic label to the patient. The second and third components together constitute a complex middle stage during which the psychiatrist perceives, classifies, summarizes and to some degree interprets statements

---

[1] The studies described here were carried out by the staff of the U.S./U.K. Diagnostic Project (Honorary Director, Dr J. Zubin), a bilateral research group based in New York at Biometrics Research, New York State Department of Mental Hygiene and the Department of Psychiatry, Columbia University; and in London at the Institute of Psychiatry (University of London). The Project psychiatrists for these studies comprised: in New York, Dr B. J. Gurland, Dr L. Sharpe, Dr T. Farkas; and in London, Dr J. E. Cooper, Dr R. E. Kendell, Dr J. R. M. Copeland, Dr N. Sartorius (attached).

made by the patient. These processes go on in the psychiatrist's mind as the interview is proceeding and are his guides in the choosing of further questions or lines of enquiry which in turn result in more information. Even from the quite general statements made up to now, it is evident that the system of information extraction and processing formed by the first three components is essentially sequential in nature. The topic at any one moment is to some extent determined by what immediately preceded it, and so divergence between alternative lines of enquiry may be cumulative as the interview proceeds.

To achieve in the final stage the high degree of comparability necessary for research purposes, all stages must be under some sort of conscious control, as far as the patient's condition and expectations of the interview allow. A complete system of standardized procedures will be needed, not merely a standardization of the final choice of a diagnostic term. I suggest that maximum agreement between a group of psychiatrists will best be achieved if account is taken of the following points:

1. *The background and training of the psychiatrists*

The task of the psychiatrists concerned will be much easier if they are all reasonably similar in general background, interests, and training. They all have to learn the same rules and common base-lines for all the ratings. They must also learn to tolerate and minimize the often inevitable discrepancy between the patient's concept and expectation of the interview and the diagnostic aims of the interviewer. If the psychiatrists have very different backgrounds and training, then a much longer period of joint training and familiarization will be necessary to reach acceptable standards of reliability.

2. *The use of detailed interview guides and rating schedules*

The same symptoms need to be covered in the interviews with all the patients and in the same manner. This is in addition to the necessarily different personal problems that every patient brings into the interview. Symptoms and behaviour elicited by such a standardized interview then need to be rated according to a set of rules and definitions. In other words, the vital details of the interviewing and rating procedures must be under control by using an itemized interview guide and rating schedule.

Mental state and history interviews with the patient and history interviews with relatives must all be included in the system.

3. *The use of diagnostic terms of accepted and explicit meaning.* A synthesis of the information from all the interviews followed by the choice of a final diagnosis can then be done with some hope of the result being repeatable, so long as conscientious use is made of a glossary of terms attached to an acceptable classification.

These points formed the basis of the system of interviews and diagnostic procedures used in the Netherne/Brooklyn and London/New York sample comparisons which will now be described.

A study of mental hospital admission statistics

Curious and even startling differences between American and British hospital admission statistics have been known for some years and particularly important are the very large and surprising differences between American and British mental hospitals in their admission rates for manic-depressive psychosis, highlighted by figures prepared by Kramer (5). In Britain, for instance, for some of the middle-aged groups of patients this diagnosis is made more than ten times as often as in the American patients of similar age. It is very unlikely that differences of this size are due entirely to the patients and the aim of this study was to discover how such differences arise.

Many influences may be responsible for this type of difference in the admission statistics and they can be divided for convenience into two groups:

(1) Real differences in the clinical states of the patients from whom the statistics originate.

(2) Differences in the psychiatric concepts and terms used by the hospital psychiatrists to describe the patients, and differences in the official collecting, coding, or recording procedures which finally produce the statistics.

One of the main aims of our study was to answer the question: 'are the differences in the official statistics really due to differences between the patients or are they due to differences between the doctors and the recording systems, or do both play a part?' The study was done in two stages, the first stage using one hospital in each city (Brooklyn State Hospital in New York and Netherne Hospital in London) and the second using a sample of admissions taken from a sufficient number of different hospitals on each side to allow the findings to be generalized to cover all mental hospital admissions in London and in New York.

Whatever clinical psychiatrists may think about the quality and usefulness of routine official admission statistics, there is no doubt they are being used more and more nowadays by administrators and planners. It is becoming increasingly important for us to know something about their nature.

The procedure adopted was to study the patients who constituted a series of consecutive admissions in each hospital, subjecting them all to standardized interviewing, rating, and diagnosis procedures, thus producing a set of diagnoses in which variations due to the psychiatrists and the data-collecting procedure are minimized. The hospital diagnoses of the same patients were also collected from the authority which produces the routine hospital statistics. The project diagnoses and ratings can be regarded as a yardstick to set against the official hospital diagnoses, so as to demonstrate the relative effects of the two main types of differences just mentioned. This gives, of course, a demonstration only of relative relationships; it is not implied that the project diagnoses are necessarily more 'correct' than any others, but it is certain that they are more uniform and that they are arrived at by repeatable methods applied by known persons.

The Brooklyn/Netherne comparison

These two hospitals were chosen because their admission statistics for the previous year had been close to the national averages: this meant that it was unlikely that they were very unusual in their catchment areas or the diagnostic habits of their doctors.

In each hospital a total of 250 admissions of patients between the ages of 20 and 59 was studied. The very young and the very old were not included because both groups introduce special problems of interviewing and diagnostic assessment. The 250 admissions were done in two stages: 145 consecutive admissions aged 35–59, which are called here the first series, and nine months later 105 consecutive admissions aged 20–34, the second series. They were split in this way for purely practical purposes and can be regarded as forming a consecutive series of 250 admissions of patients aged 20–59 at each hospital between these age limits over a period of roughly three months.

Interviewing and diagnostic procedures

The interviewing and diagnostic methods used in this study have been described elsewhere (2) and only a brief account will be given here.

As soon as possible after admission, usually within 48 hours, every patient was interviewed by one of the project psychiatrists, using a standardized mental state schedule. During the next few days, or as soon as the patient was well enough, he was interviewed again by the same psychiatrist, using a standardized history schedule put together specially for this study. The nearest available relative of the patient was also interviewed, usually by a project social worker but sometimes by the original project psychiatrist, using the same standard history schedule. If necessary, the social worker visited the relative at home to obtain this second history interview. The two versions of the history schedule were then combined into a final single set of ratings after discussion between the two interviewers.

Using the information from these three interviews, the project psychiatrist who had done the mental state and the history interviews then made the final project diagnosis using the eighth edition of the *ICD* and the accompanying 'Glossary of Mental Disorders' compiled for use in the U.K. by the General Register Office (3).

The standardized mental state interview referred to above consisted of the seventh edition of the Present State Examination (PSE), modified for this study by the addition of some 200 items from the 'Psychiatric Status Schedule'.

The PSE was developed originally at the Maudsley Hospital by Dr John Wing and his colleagues (11). In addition to its use in this project, the PSE is being used by the WHO in their

International Pilot Study of Schizophrenia, so it is now being widely used on an international scale. The interview schedule consists of approximately 500 items which are rated as the interview proceeds. The ratings do not necessarily depend upon the reply of the patient, but represent the interviewer's clinical judgement as to whether that particular abnormality is present. A question is provided for each rating and it is usually possible for the interviewer to use this, but he is also allowed to ask additional questions at his own discretion. Thus this is not a questionary but a clinical interview which has been standardized as much as possible. As might be expected, a good deal of instruction and practice are necessary before a standardized interview of this detail and complexity can be used with comfort and reliability, but there is no doubt that this can be achieved. Evidence that satisfactory standards of inter-rater reliability and repeatability can be achieved has already been published (4).

For the purposes of this hospital study, this interview schedule was combined in a single document with approximately 200 items of the Mental Status Schedule developed by Spitzer *et al.* (9). The ground covered by these two standardized interviews is necessarily very similar, so it was possible to use this combined schedule without much difficulty and to complete the whole interview in about 75 minutes. The presence of these extra items allowed the use of a modified form of Spitzer and Endicott's (10) 'Diagno 1' computer program to produce a diagnostic breakdown for both series of patients which is necessarily free from inter-psychiatrist variation. This is a valuable check on the consistency of the project diagnosis.

It was decided to use for the 'project diagnosis' the new eighth edition of the *ICD* which from this year onwards will be widely used in many countries. A glossary of terms or operational definitions has been produced for this classification by a committee in the U.K. (3), similar to that in the American Psychiatric Association (1968) *Diagnostic and Statistical Manual* (1). This glossary contains only descriptions of symptoms and abnormal behaviour, and so these diagnoses are 'descriptive' or 'operational' definitions; they are not aetiological diagnoses and do not utilize interpretative statements about psychological defence mechanisms.

The project psychiatrist dealing with the admission made a diagnosis from the international classification as soon as all the interviews had been completed. This diagnosis was then discussed with one other project psychiatrist before being accepted as the project final diagnosis. If the case was a difficult diagnostic problem, or if the two psychiatrists disagreed, then it was discussed in detail with a third project psychiatrist before arriving at a final diagnosis. In other words, the project final diagnosis is an agreed consensus diagnosis in which considerable effort has been made to base the decision on a given set of descriptions. In the following discussion, this will be contrasted with the official hospital diagnoses for these same patients.

In the British series, the official diagnosis which appears in the national statistics was obtained from the Ministry of Health. For the American series, the official diagnosis was collected from the New York State Department of Mental Hygiene, Albany, after it had passed through the usual channels.

None of the information gathered about these patients by the project staff was transmitted to the hospital psychiatrist or the nursing staff and cases were not discussed with them, even informally. Neither were the patients' hospital case-notes examined by the project staff. Thus there was no contamination of the hospital diagnosis by the project data, or vice versa.

The Brooklyn/Netherne results

If we regard the more uniform project diagnoses as a base-line, comparison with the hospital diagnoses for each hospital in turn will indicate the relative degree of diagnostic bias or distortion in the official diagnoses of the same admissions. The second comparison to be made is between the project diagnoses for the two series of patients, which is the best way we have of detecting any differences in the clinical states of the two series of patients.

Table 3.2 shows the project and hospital diagnoses of the first series of 145 consecutive admissions aged 35–59 years at Netherne and Brooklyn. Seven categories are used: all varieties of schizophrenia, including 'schizo-affective psychosis' are put together; the term 'affective' includes all varieties of manic-depressive psychosis, other depressive psychosis, and also depressive neurosis. The depressive neuroses were included here

TABLE 3.2. *Brooklyn and Netherne: first series of consecutive admissions (aged 35–59 years). Hospital diagnosis (Albany and Ministry of Health) and project diagnosis. Percentage table*

	Brooklyn				*Netherne*		
	Hospital		*Project*		*Project*		*Hospital*
				n = 145			
Schizophrenia	56·6	*	29·7		22·8	*	35·2
Affective	16·6	*	36·6	*	58·6	†	46·2
Neuroses	0·7	*	6·9		2·8		4·8
Personality	0·7		1·4		0·7		1·4
Alcoholism	19·3		16·6	*	9·0		7·6
Organic	2·8		3·4		2·1		0·7
Other	3·4		5·6		4·1		4·1
Total	100		100		100		100

* p = 0·01 for significance of difference between the figures on either side, using Critical Ratio.
† p = 0·05.

because of the considerable practical difficulties in making a reliable differentiation between depressive psychosis and neurosis. In any case, the number of depressive neuroses is not large and whether they are included here or not makes no difference to the major findings. The other terms, 'other neuroses', 'personality disorders', 'alcoholism', 'organic disorders', or 'other', are self-explanatory. The great majority of cases in all the tables fall within the first two groups.

At each hospital the project made fewer diagnoses of schizophrenia on the same patients than the hospital, but more so at Brooklyn than at Netherne (at Brooklyn 56·6 per cent hospital against 29·7 per cent project, and at Netherne 35·2 per cent against 22·8 per cent project). For affective disorders, the reverse is true—the project made more diagnoses of this type than either hospital and again more so at Brooklyn than at Netherne (Brooklyn 16·6 per cent hospital against 36·6 per cent and at Netherne 46·2 per cent hospital against 58·6 per cent project). The project breakdowns, which form the two central columns in all these tables, show that these two groups of patients are by no means as dissimilar as the hospital breakdown in the two outer columns suggests. A very significant difference of 21 per cent for schizophrenia (56·6–35·2 per cent between the two sets of hospital figures is reduced to an insignificant difference on the project figures of 7 per cent (29·7–22·8 per cent). For affective disorders a very signifficant difference of 30 per cent between the hospital figures (46·2–16·6 per cent) is reduced to a still significant but smaller difference of 22 per cent (58·6–36·6 per cent).

Table 3.3 shows the second series of younger patients at the same two hospitals and both these effects are present to an even greater degree at Brooklyn, but at Netherne these differences have diminished to an insignificant level. At Brooklyn the hospital made a diagnosis of schizophrenia in 77·1 per cent of the admissions, the project in only 36·2 per cent. The diagnosis of affective disorder in these younger patients was virtually not made by the hospital at Brooklyn, whereas the project used this term for 36·2 per cent of the same patients. At Netherne, the hospital breakdown is very close to that of the project.

TABLE 3.3. *Brooklyn and Netherne: second series 105 admissions (aged 20–34 years). Hospital and project diagnoses. Percentage table*

	Brooklyn				*Netherne*	
	Hospital		*Project*		*Project*	*Hospital*
				n = 105		
Schizophrenia	77·1	*	36·2		30·5	32·4
Affective	1·9	*	36·2		31·4	27·6
Neuroses	—		1·0	*	16·2	12·4
Personality	1·0		4·8		9·5	17·1
Alcoholism	2·9		4·8		2·9	—
Drugs	2·9		7·6		3·8	4·8
Organic	2·9		3·8		1·0	3·8
Other	11·4		5·7		4·8	1·9
Total	100		100		100	100

* p = 0·001 for significance of difference between the figures on either side, using Critical Ratio.

For this age-group of patients, the project diagnosis shows that the large differences between the two sets of hospital diagnoses are almost entirely due to the psychiatrists and the collecting systems, since the differences between the project figures in the centre columns in this table are small and do not quite reach a significant level, except for the neuroses.

Table 3.4 shows the same breakdown for both the series together, a total at each hospital of 250 admissions of patients aged 20–59 years. The differences we have just discussed are all present in this summary table. They can perhaps best be summarized by saying that the Netherne Hospital diagnostic system for both schizophrenia and affective disorders is intermediate between the project and the Brooklyn system. The project is the most prone to diagnose affective disorder and Brooklyn the least. Brooklyn is the most prone to diagnose schizophrenia and the project the least.

TABLE 3.4. *Brooklyn and Netherne: both series 250 admissions (aged 20–59 years). Hospital and project diagnoses: Percentage table*

	Brooklyn				*Netherne*		
	Hospital		*Project*		*Project*		*Hospital*
			n = 250				
Schizophrenia	65·2	†	32·4		26·0	*	34·0
Affective	10·4	†	36·4	*	47·2	*	38·4
Neuroses	0·4		4·4		8·4		8·0
Personality	0·8		3·2		4·4		8·0
Alcoholism	12·4		11·6		6·4		4·4
Drugs	1·2		3·2		2·8		2·0
Organic	2·8		3·6		1·6		2·0
Other	6·8		5·2		3·2		3·2
Total	100		100		100		100

* $p < 0{\cdot}05$ for significance of difference between the figures on either side using Critical Ratio.

† $p < 0{\cdot}01$.

The most striking thing about this table is that on the hospital breakdowns (the two outside columns) these two groups of admissions are significantly different for the first five of the diagnostic categories, particularly so for schizophrenia and affective disorders. On the project figures, however, only affective disorders remain significantly different, although the figures for schizophrenia are nearly so.

We can now answer the question asked at the beginning of this paper, at any rate for these two hospitals. Almost all of the large differences between the hospital admission statistics of these two hospitals are artefacts due to differences in the diagnostic practices of the psychiatrists and differences in the recording systems. The only real difference in the patients seems to be that rather more patients with affective disorders, that is mainly depressive illnesses, are entering Netherne than are entering Brooklyn.

The sampling studies: method

Netherne and Brooklyn were chosen because their admission statistics were known not to be unusual, so it is very likely that these findings reflect to a considerable extent the position at many other similar hospitals. The implications in these results are important, however, and it was thought worthwhile to do a similar study on a sample of admissions to other hospitals in London and New York that would show to what extent our conclusions could be generalized with confidence.

There are nine state hospitals which between them serve the whole of metropolitan New York, so these seemed to suggest themselves as the area and the hospitals to be used. It turned out that the number required for the total sample (about 180) was very close to half the number of patients expected to be admitted to all the nine hospitals together during an 'average week' (based upon the previous year's figures). Therefore, a number of patients at each hospital equal to half of that hospital's average number of weekly admissions was selected, spread over seven consecutive days.

British area mental hospitals are smaller and hence more numerous than American state hospitals and the Greater London area is served by well over twenty mental hospitals. It was necessary to study only a sample of these and, after a detailed study of the hospital catchment areas in and around London, it was finally decided to confine the study to those area mental hospitals who had at least 50 per cent of their catchment population living within the boundary of the new Greater London Council. Eighteen hospitals fell into this category and it was decided to take a sample of half of these. Type of catchment area was also taken into consideration, since it is well known that the type of catchment area serving a hospital has an important effect on the admission diagnoses: hospitals serving the central broken-down and crowded areas of a large city usually have a higher proportion of schizophrenics amongst their patients than their rural counterparts. This was allowed for in the selection process by studying the position of the hospital catchment areas on a large-scale map of London, and assigning them all to one of three groups—catchment areas all or mainly central, catchment area mixed, and catchment area all or mainly peripheral. Six hospitals fell into each group and so three needed to be picked from each group. A procedure based on random numbers was devised and a weighting procedure was included which gave the hospitals with the largest admission rates for the previous year more chance of being picked than those with the smallest. Each of the nine selected hospitals was visited in turn and a series of consecutive admissions of patients between the ages of 20 and 59 years was studied; the number of admissions at each hospital being proportionate to the total number of admissions for the previous year. The final total sample size chosen was 174 admissions, which meant about 20 admissions at each hospital.

Only one change of any importance was made in the interviewing procedures that were used at Brooklyn and Netherne. Because of the extra travelling involved in visiting nine hospitals, it was not practicable to rely upon the interviewing of relatives as well as patients: in the Brooklyn/Netherne study this part of the interviewing had proved to be the most time-consuming and difficult to arrange of all. Somewhat against our expectations, when the contribution of the interview with the relatives towards the final project diagnosis was studied, it was found to be not very large. In over 80 per cent of cases, the provisional diagnosis made at the end of the mental state interview was in the same major category as the final diagnosis for all the interviews which had been used. Furthermore, the small number of major changes between schizophrenia and affective psychosis tended to cancel each other out in the simple type of diagnostic breakdown presented in the tables.

This meant that a great deal of time and effort could be saved by omitting the interview with the relative, with an effect upon diagnostic accuracy which would be very small compared to the very large differences that were likely to be found between the hospital and the project diagnoses. Accordingly, relatives of patients were not interviewed in the sampling studies. The mental state interview and the diagnostic procedures with the eighth revision of the *ICD* and its accompanying British glossary were the same as for the Brooklyn/Netherne study, except for the inclusion of more of Spitzer's Diagno items in the mental state interview.

The sampling studies: results

The extent to which the diagnostic breakdowns of the Netherne and Brooklyn series correctly reflect their respective sample surveys must now be considered.

TABLE 3.5. *London sample, and Netherne (both series). Project and hospital diagnoses. Percentage table*

	Netherne				*London*	
	Hospital		*Project*		*Project*	*Hospital*
	n = 250				n = 174	
Schizophrenia	34·0	*	26·0	*	35·0	33·9
Affective	38·4	*	47·2		43·7	39·1
Neurosis	8·0		8·4		4·0	5·7
Personality	8·0		4·4		2·9	4·6
Alcoholism	4·4		6·4		4·6	3·4
Drugs	2·0		2·8		0·6	0·6
Organic	2·0		1·6		3·4	1·7
Other	3·2		3·2		5·7	10·9
Total	100		100		100	100

* $p < 0{\cdot}05$ for significance of difference between the figures on either side, using Critical Ratio.

Table 3.5 shows the diagnostic breakdown for the Netherne series of 250 admissions and for the London sample from nine area mental hospitals. Looking first at the two sets of official hospital diagnoses in the two outer columns, it is evident that the figures are virtually identical, so from the point of view of Ministry of Health statistics the Netherne patients were very typical of the whole of London. The project versions of these two sets of admissions are not so similar, however, and there are significantly fewer schizophrenics entering Netherne—only 26·0 per cent—than in the sample survey—35·0 per cent. This means that the Netherne psychiatrists are more prone to diagnose schizophrenia than the psychiatrists in the other London hospitals; they have, so to speak, an American tendency in this direction. In this table, the project diagnoses are being used to show up dissimilarities which are not apparent from the hospital diagnoses in contrast to the previous tables, in which the project diagnosis always tended to reduce apparent differences. The project figures are very close to the sample hospital figures in all the categories, which is encouraging in that it suggests that the procedures for producing a project diagnosis must be reasonably similar to the diagnostic habits of many other British hospital psychiatrists. It is worth noting, however, that the fairly crude arrangement of data being used in these tables can hide a good many individual disagreements which often cancel each other out in this method of presentation.

We may conclude that the Netherne series mirrors the London sample exactly according to official statistics, but that a lack of schizophrenic admissions is in fact hidden by the slight bias towards diagnosing schizophrenia possessed by the Netherne psychiatrists. In other respects, the Netherne series is reasonably similar to the London sample.

Table 3.6 shows the same figures for the Brooklyn series and

TABLE 3.6. *New York sample, and Brooklyn (both series). Project and hospital diagnoses. Percentage table*

	Brooklyn				*New York*		
	Hospital		*Project*		*Project*		*Hospital*
Schizophrenia	65·2	*	32·4		29·2	*	61·5
Affective	10·4	*	36·4		32·3	*	6·8
Neurosis	0·4		4·4		1·6		2·6
Personality	0·8		3·2		4·2		1·0
Alcoholism	12·4		11·6	*	22·9		19·8
Drugs	1·2		3·2		3·1		—
Organic	2·8		3·6		2·6		5·2
Other	6·8		5·2		4·2		3·1
Total	100		100		100		100

* $p < 0·01$ for significance of difference between the figures on either side, using Critical Ratio.

the New York sample. For all except alcoholism, both outer and inner pairs of columns in this table match each other very closely. Brooklyn is unusual in that it takes only about half as many alcoholics as are in the sample. The psychiatrists and hospital systems in the other state hospitals differ from the project in their diagnostic habits in the same direction and to about the same and very striking extent as did those at Brooklyn.

We can safely conclude that our major findings about the directions and the size of the diagnostic biases in the hospital statistics noted on the Netherne/Brooklyn series can be generalized to the rest of the similar hospitals in London and New York, and quite probably to many other cities, hospitals, and psychiatrists elsewhere in the two countries.

Consistency of the project diagnosis

The crucial issue of the consistency of the project diagnosis on either side of the Atlantic will now be examined in more detail. Three ways are available for doing this:

(1) An examination of the diagnoses made by the project staff when they exchanged between the two sides of the study.

(2) A comparison of the project diagnoses with the diagnoses made from some of the interview items by Diagno 1, a computer program (10).

(3) A comparison of the project diagnoses with the diagnoses derived from a canonical variate analysis.

1. EXCHANGE OF PROJECT STAFF

The reliability and repeatability of the PSE have been studied and found to be satisfactory (4, 11) but these studies were done when the psychiatrists were in close and continual contact with each other. However carefully definitions or ratings are initially learned and used, divergence between raters once they have stopped regular joint interviewing and rating sessions is likely, so it was felt advisable to take other precautions as well as to keep variation down to a minimum. Because of the need for training in the interviewing and rating techniques, there was a considerable interchange of project psychiatrists between the sides. One of the American team spent one month at Netherne and all three members of the British team spent one month at Brooklyn. As a result of this, a third of the Brooklyn admissions in the first series of 145 were interviewed by two of the British team, which meant that there was a constant exchange of ideas and information about the diagnostic classification in use as well as about the more detailed interviewing and rating procedures.

This exchange of psychiatrists also makes possible a useful check on the consistency of the project diagnostic procedure between the two hospitals. The British half of the team interviewed and made the provisional diagnosis for 49 of the Brooklyn admissions, so the distribution of diagnoses in these 49 patients can be compared with that of the remaining 96 dealt with by the New York team. Table 3.7 shows the distribution of diagnoses in these two groups and it is evident that they are very similar. Such differences as exist between the two distributions are in the opposite direction to the differences between the two sets of hospital and project diagnoses being discussed.

TABLE 3.7. *Diagnoses given by project to Brooklyn sample, by team membership of interviewer. Percentage table*

Diagnosis	*London team* (n = 49)	*New York team* (n = 96)
Manic	4·1	8·3
Depressed	28·6	30·2
Schizophrenic	32·7	28·2
Neurosis, pers. dis.	10·2	8·3
Alcoholic	16·3	16·7
Other	8·1	8·3
Total	100·0	100·0

2. DIAGNO I

The presence of the 200 items from the Psychiatric Status Schedule in our combined interview schedule enables us to use a slightly modified form of Spitzer and Endicott's 'Diagno 1' (10) to see if the project diagnoses bear the same relationship to the patients' symptoms in each hospital. 'Diagno 1' carried out a logical decision-tree analysis of Mental Status Schedule items marked as present and arrived at a diagnosis on that basis. Since all the patients are subjected to the same set of decisions, the Diagno 1 diagnostic breakdown for different groups of patients can be regarded as a consistent base-line against which to check the consistency of other methods of arriving at a diagnosis; it can be used to examine the relative bias of the project diagnosis just as that has been used to examine the relative bias of the hospital diagnoses.

The Diagno 1/project ratio at Brooklyn should be similar to that at Netherne if the two project teams have been using the diagnostic procedure consistently, at least for the proportion of diagnoses in the two major categories of schizophrenia and affective illness. Results from Diagno are available for the first two series of 145 admissions each to Netherne and Brooklyn and also from the London and New York samples (see Tables 3.8 and 3.9).

TABLE 3.8. *Project and Diagno 1 diagnoses for Brooklyn and Netherne first series (aged 35–59 years). Percentage table*

	Brooklyn n = 145 1 *Diagno*	 2 *Project*	*Netherne* n = 145 1 *Diagno*	 2 *Project*
Schizophrenia	46	30	39	23
Affective illness	19	37	26	59
Other	35	33	35	18
Total	100	100	100	100

Schizophrenia

Diagno/project ratio at Brooklyn $= \frac{46}{30} = 1{\cdot}53$

Diagno/project ratio at Netherne $= \frac{39}{23} = 1{\cdot}70$

Affective illness

Diagno/project ratio at Brooklyn $= \frac{19}{37} = 0{\cdot}51$

Diagno/project ratio at Netherne $= \frac{26}{59} = 0{\cdot}44$

TABLE 3.9. *New York and London samples: project and Diagno 1 diagnoses. Percentage table*

	New York			London		
	Project	*Diagno*	*Diff.*	*Project*	*Diagno*	*Diff.*
		n = 192			n=174	
Schizophrenia	29	45	16	35	45	10
Affective	32	14	18	44	28	16
Alcoholic	23	18	5	5	5	0
Other	16	23	5	17	23	6

Schizophrenia

$$\text{Diagno/project ratio in London} = \frac{45}{35} = 1{\cdot}29$$

$$\text{Diagno/project ratio in New York} = \frac{45}{29} = 1{\cdot}55$$

Affective illness

$$\text{Diagno/project ratio in London} = \frac{28}{44} = 0{\cdot}64$$

$$\text{Diagno/project ratio in New York} = \frac{14}{32} = 0{\cdot}44$$

There are several ways of examining the differences between two pairs of diagnostic breakdowns, but a suitable method for our purposes is to work out, for each side of the study separately, the ratio of Diagno to project diagnoses for schizophrenia and affective illnesses. If the two halves of the project used their system in similar fashion, the relationship between the Diagno and the project diagnostic breakdowns for the major categories should be similar on the two sides.

These comparisons are given for the first part of the Netherne/Brooklyn series in Table 3.8. For schizophrenia, the Diagno/project ratio was 1·53:1 at Brooklyn and 1·70:1 at Netherne. For affective illnesses the Brooklyn ratio was 0·51:1 and the Netherne ratio 0·44:1. The same ratios for the London/New York sampling study are given in Table 3.9. In all instances, the members of these pairs of ratios seem similar enough for it to be concluded that both sides of the project diagnosis had a very similar relationship to the uniform Diagno. There is a small difference between these two series which is worth noting. In the sampling study, the ratios changed a little in the direction suggesting that the London team had become a little more similar to Diagno than the New York team. This is most easily seen in the figures for schizophrenia in Table 3.9, where the London side are 10 per cent below Diagno (45–35 per cent) whereas the New York side are 16 per cent below (45–29 per cent). This difference is not significant but it is worthy of comment since a similar tendency can be seen in the comparison of the project diagnoses with the canonical variate analysis.

There was less interchange between the project staff than in the first series, so this difference may be due to genuine diagnostic variations between the two sides, although we were not subjectively aware of this when meeting and discussing the diagnoses. An alternative or additional explanation is that there were more patients in the New York series with whom the interviews were difficult due to language problems or poor cooperation. This makes it less likely that the key svmptoms needed for a diagnosis of schizophrenia would be elicited or detected, but such difficulties would not penalize to the same extent the severe depressive symptoms, some of which are more a matter of observation than communication.

Fortunately this difference is small and it is quite insignificant when compared with the very large differences between the London and New York hospital diagnoses that were the object of our inquiry.

3. CANONICAL VARIATE ANALYSIS

For other purposes, a canonical variate analysis has been carried out upon the ratings from the mental state interviews of those patients in the Brooklyn/Netherne series with a diagnosis of schizophrenia or affective illnesses. This analysis is also of interest as a check upon the project diagnosis, since it can be used to distinguish between two populations, for instance, schizophrenics and non-schizophrenics, on the project diagnosis, in such a way as to minimize overlap between the two, and using for this the 'average' diagnostic criteria of the several contributors to the total number of project diagnoses. The diagnostic breakdown of a group of patients according to this canonical variate will probably not be identical to that produced on the same patients by any one individual of the project staff and the differences will be caused by the individual's variation in relation to his colleagues. Similarly, the over-all project breakdown will not be quite the same as that produced by the canonical variate, which will tend to iron out the individual peculiarities of the project staff. These differences are shown in Table 3.10 and, as with Diagno, project staff in London make proportionately rather more diagnoses of schizophrenia than the project staff in New York. The same ratios are used as for Diagno 1. Again, however, the differences between the two sides of the project are very small in comparison to the large differences between the hospital and the project in New York.

TABLE 3.10. *Comparison of project and canonical variate diagnoses. Percentage table*

	Admissions in U.S.		*Admissions in U.K.*	
	1 *Project*	2 *Can. Var.*	1 *Project*	2 *Can. Var.*
Schizophrenia	44·3	50·8	31·9	32·4
Psychotic affective	36·6	33·9	38·7	40·2
Neurotic	19·1	15·3	29·4	27·5
Total	100	100	100	100

Schizophrenia

$$\text{Canonical variate/project in U.S.} = \frac{50\cdot8}{44\cdot3} = 1\cdot146$$

$$\text{Canonical variate/project in U.K.} = \frac{32\cdot4}{31\cdot9} = 1\cdot015$$

Psychotic affective

$$\text{Canonical variate/project in U.S.} = \frac{40\cdot2}{38\cdot7} = 1\cdot039$$

$$\text{Canonical variate/project in U.K.} = \frac{33\cdot9}{36\cdot6} = 0\cdot926$$

4. DISCUSSION

a. The findings of the investigations

It is clear from the similarity between the project and the hospital diagnoses on the London sample (Table 3.5) that the project diagnostic system is not particularly eccentric in its biases, at any rate when the results are tabulated in the fairly crude form used here for discussion of large-scale official statistics. The project diagnoses agree with the combined diagnoses of many different London psychiatrists to a greater extent than we had anticipated. There are, however, still many disagreements in individual cases, but these tend to cancel each other out in the fairly crude over-all percentage lists presented here.

The really startling disagreements between the project and U.S. hospital diagnoses show how difficult it is for British and American psychiatrists to make use of each other's statistics, research reports, or clinical descriptions which do not give quite explicit details about the diagnostic criteria used. The use of the routinely collected hospital diagnoses for assessment of response to treatment or prognosis will produce results of only local interest. It is evident that very different meanings are attributed to the familiar terms of 'schizophrenia' and 'manic-depressive psychosis', and, in particular, these findings show that the concept of schizophrenia being used by the psychiatrists in the New York state hospitals must be much broader than that of their London counterparts, and very much broader than that of the project diagnostic system based upon the eighth *ICD*. It has been known for many years that differences of this type existed between American and British psychiatrists, but the knowledge was of a personal or anecdotal type. This use of a standardized diagnostic system on a large scale in a mental hospital setting shows how large an effect such differences can have upon hospital admission figures which must bear a very close relationship to the respective national statistics.

The project diagnoses show that as far as simple diagnostic labels are concerned the population of admissions to mental hospitals in London is fairly similar to state mental hospitals in New York, except for alcoholism. (It must be remembered that the New York City receiving hospitals were not included in the study, since they do not contribute towards the official statistical returns for the state.) In the first series of patients at Brooklyn and Netherne (aged 35–59) there was, however, one significant difference of some interest on the project diagnoses—more patients with depressive disorders entered Netherne than Brooklyn. A closer look at their symptoms shows that the

British patients with depression were not only more numerous than the American, but that their symptoms were also more severe; the reverse might have been expected in view of the greater ease of entry to Netherne as a voluntary or 'informal' patient. It must be emphasized that these data do not allow any statements about the incidence or prevalence of these illnesses in the general population of the two cities; the study was not designed for this purpose.

b. The use of the standardized diagnostic system

Fears are often expressed about the inhibiting effect of a standardized interview upon the interviewer and the patient. This is a reasonable fear, but experience shows that with training and practice it can be reduced to a level at which interviewer and patient only rarely appear uncomfortable. An independent observer is not usually irritated so long as he remembers that this is a research procedure with a specific object in mind and is not just a slightly modified form of an ordinary clinic or ward interview. A standardized interview only finds what it is designed to find, but this is no disadvantage so long as this is acknowledged and the procedure used where this restriction is not important. After using this whole system of interviews and final diagnostic choice procedure on a large scale for a period of roughly two years, a strong subjective impression is left that the most unreliable part of the whole process is the final one—the use of a glossary as a guide to the final choice of a diagnostic term. The present eighth *ICD* glossary produced in the U.K. (3) is a major step forward but it still represents only a guide rather than a set of definite instructions. This also applies to the American *DSM*, second edition (1). For research purposes, it seems essential to try and make the final choice of a diagnostic term by means of a standard set of decisions; this is most conveniently done by using computer programs which work from the items and symptoms directly. Diagno 1, used in this investigation, was of great value as a check on the project diagnoses, even though it had a considerable American bias, being written by an American psychiatrist.

With a good deal of time, trouble, and programming, any set of decisions made by a psychiatrist can be turned into a computer program and, in the absence of objective laboratory tests and measures, psychiatric research will doubtless come to rely more and more heavily upon such aids.

The note elsewhere in this volume on the Catego program (12), indicates the method used to produce a computer program which will give a diagnosis based upon the items of the PSE, so removing the inter-psychiatrist variation from this most vulnerable final diagnostic choice. This is the last stage in the standardization of our diagnostic system—a system which has aimed to remove as many as possible of the sources of variation while interfering as little as possible with the clinical interview.

REFERENCES

1. American Psychiatric Association (1968). *Diagnostic and Statistical Manual, Mental Disorders*, 2nd edn. (Washington D.C.).
2. Cooper, J. E., Kendell, R. E., Gurland, B. J., and Sartorius, N. (1969). 'Cross-national study of diagnosis of mental disorders: some results from the first comparative investigation', *Amer. J. Psychiat.* **125**, Suppl. (April issue).
3. General Register Office (1968). *Studies on Medical and Population Subjects, No. 22*, 'A Glossary of Mental Disorders' (London: HMSO).
4. Kendell, R. E., Everett, B., Cooper, J. E., Sartorius, N., and David, M. E. (1968). 'The reliability of the Present State Examination', *Social Psychiat.* **3**, 123–9.
5. Kramer, M. (1961). *Proceedings Third World Congress of Psychiatry* (*Montreal*), **3**, 153–60.
6. —— (1969). 'Cross-national study of diagnosis of mental disorders: origin of the problem', *Amer. J. Psychiat.* **125**, Suppl. (April issue).
7. Kreitman, N. (1961). 'The reliability of psychiatric diagnosis', *J. ment. Sci.* **107**, 876–86.
8. Shepherd, M., Brooke, E. M., Cooper, J. E., and Lin, T. (1968). 'An experimental approach to psychiatric diagnosis', *Acta psychiat. scand.*, Suppl. 201.
9. Spitzer, R. L., Fleiss, J. L., Burdock, E. I., and Nardesty, A. S. (1964). 'The mental status schedule: rationale, reliability and validity', *Comprehensive Psychiat.* **5**, 384–95.
10. —— and Endicott, J. (1967). 'Diagno: a computer program for psychiatric diagnosis utilizing the differential diagnostic procedure', *Arch. gen. Psychiat.* **183**, 746–56.
11. Wing, J. K., Birley, J. L. T., Cooper, J. E., Graham, P., and Isaacs, A. D. (1967). 'The reliability of a procedure for measuring and classifying Present Psychiatric State', *Brit. J. Psychiat.* **113**, 499–515.
12. —— (1970). 'A standard form of psychiatric present state examination and a method for standardizing the classification of symptoms' (this book, pp. 93–108).
13. Zubin, J. (1967). 'Classification of the behaviour disorders', *Ann. Rev. Psychol.* **18**, 373–401.
14. —— (1969). 'Cross-national study of diagnosis of mental disorders: methodology and planning', *Amer. J. Psychiat.* **125**, Suppl. (April issue).

Reliability of psychiatric assessment in international research

from the WHO international pilot study of schizophrenia

N. SARTORIUS, EILEEN M. BROOKE, and T. LIN
Mental Health Unit, World Health Organization, Geneva, Switzerland

Introduction

Reliability in case-finding is a primary essential for epidemiological studies. In the past few years WHO has conducted an extensive programme designed to prepare the ground for future international studies in psychiatric epidemiology and since 1967 has conducted a pilot study as a first step to this end (1). As schizophrenia is an illness which is commonly diagnosed in all countries and accounts for a large proportion of psychiatric morbidity, it was selected for the pilot study. The aim of the International Pilot Study of Schizophrenia (IPSS) is to answer the question whether comparable cases of schizophrenia can be found in various cultures by psychiatrists differing in nationality, language, and school of training. Field research centres in nine countries are participating in the IPSS: Aarhus (Denmark), Agra (India), Cali (Colombia), Ibadan (Nigeria), London (U.K.), Moscow (U.S.S.R.), Taipei (Taiwan), Prague (Czechoslovakia).

In order to achieve reliability in case-finding, three main conditions require to be satisfied:

(*a*) When two or more psychiatrists are observing the same patient, they should agree about the symptomatology and behaviour;

(*b*) When simultaneously observing the same patient, they should agree upon the diagnosis;

(*c*) Patients for whom the same diagnosis is made should have similar symptomatology and behaviour.

This paper summarizes the results of some of the work which has been done to try to ensure reliability of assessment of the mental status of schizophrenic patients, in relation to (*a*) and (*b*) of the criteria stated. More detailed results, together with a discussion of what has been done in regard to condition (*c*), will be presented elsewhere.

Selection of patients for inclusion in the study

The procedure for selecting the patients was standardized, the object being to include, in the first instance, only patients with functional psychoses. The selection was first restricted to patients who contacted specified psychiatric facilities in the nine centres, who were aged 15–44 (thus largely excluding those with psychiatric disorders of the senium), and who had resided in a defined catchment area for at least six months, as this could be taken to indicate some stability of residence in the area, which would be useful in a follow-up study. Patients satisfying these criteria were examined by a psychiatrist who completed a screen form which contained operational criteria for the exclusion of patients showing chronicity, or the presence of organic causative or concurrent conditions. From the residue, those patients with symptoms indicative of functional psychosis were chosen for inclusion in the study.

Use of standardized research instruments and procedures

In order to further uniformity in the examination of the patients, all centres are using the same research instruments and standardized procedures, which have had to satisfy several conditions:

(*a*) They had to be applicable to patients from various cultures and be capable of use by psychiatrists with different backgrounds and training.

(*b*) They should produce maximum reliability in the psychiatric assessment of patients.

(*c*) Because of their use by several psychiatrists in each centre in assessing a large number of patients, they had to be as simple as possible and applicable by the collaborating investigators after a short and easy training.

In seven of the centres it was necessary to translate the research instruments into one or more local languages. How to ensure the reliability of answers to questions which have been translated into different languages would, in itself, provide an independent research project. In the IPSS, the centres were asked to assess the value of their translations by making a back-translation into English, comparing the results with the original version and using the discrepancies detected in this way to improve translations.

Apart from the PSE schedule described earlier, slightly modified for the specific requirements of the WHO study, there are also schedules concerned with the psychiatric history, social description, and diagnostic assessment.

Preliminary training in the use of the PSE schedule

At the beginning of the study the collaborating investigators took part in a seminar organized by the Medical Research Council Social Psychiatry Unit in London. During this seminar they were trained in the use of the PSE schedule, by first observing and then taking an active part in a number of PSE interviews. Psychiatrists who were well acquainted with the schedule worked with small groups of psychiatrists from the centres. Video-taped interviews in which the schedule was used were viewed, rated, and discussed, special attention being paid to items where the raters disagreed. The psychiatrists then returned to their centres and phase 1 of the project began.

Reliability testing in phase 1 of the IPSS

During this phase a number of simultaneous interviews were carried out in each centre, in which one psychiatrist asked the question in the PSE schedule, and both he and a second, observing, psychiatrist rated the responses. By comparing the pairs of ratings of psychiatrists presented with the same material, the extent of their agreement or disagreement was assessed, and it was possible to detect whether in any centre, one psychiatrist tended to rate consistently higher or lower than another, either throughout the schedule or in certain areas of psychopathology.

Reliability between psychiatrists from different centres was assessed during exchanges of visits in London and Washington, and towards the end of this phase a film of a PSE interview was circulated to seven of the nine centres and rated by fifteen collaborating psychiatrists. The diagnoses given in this case are shown in Table 3.11.

TABLE 3.11. *Diagnostic assessment made from filmed interview of a London patient, rated in seven centres by fifteen collaborating psychiatrists*

Field research centre	*Psychiatrist's identification number*	*Diagnosis and category in the* ICD (*eighth revision*)
Aarhus	012	295.9 Schizophrenia unspecified
	013	781.9 Hallucinations and delusions
Agra	021	295.9 Schizophrenia unspecified
	022	295.3 Schizophrenia paranoid
	023	295.9 Schizophrenia unspecified
Cali	031	295.9 Schizophrenic reaction (chronic)
	032	295.9 Schizophrenia chronic
Ibadan	041	295.8 Schizophrenia-like psychosis
	042	295.7 Schizophrenia with depressive features
	043	295.7 Schizophrenia schizo-affective
Taipei	071	295.3 Schizophrenia paranoid
	072	295.3 Schizophrenia paranoid
Washington	081	295.9 Schizophrenia chronic
	082	295.9 Schizophrenia chronic
Prague	091	295.3 Schizophrenia paranoid

As can be seen from the table, there was good agreement on a general diagnosis of schizophrenia, only one psychiatrist being undecided as to whether the patient was psychotic or not. But agreement on a specific form of schizophrenia was much lower. In most cases the psychiatrists made a firm diagnosis, in spite of the fact that they did not know much about the patient's history or background. This fact was mentioned by

Wing (2) and will be referred to later. It is, however, only fair to point out that the majority of psychiatrists represented in Table 3.11 had had previous experience of psychiatry in either the U.K. or the U.S. Later on it will be possible to provide films made in other centres, such as Moscow and Prague, for multiple ratings.

Changes in the research instruments preparatory to phase 2

Although the results of comparing the ratings in simultaneous interviews done in the centres showed an encouraging measure of agreement, several changes in the instruments and procedures were made as a result of the experience gained in this preliminary phase. Some questions in the instruments were made more precise, in others the ratings were more clearly defined. Perhaps one of the more important steps taken was the production of a glossary of psychopathological terms and psychiatric diagnoses for use in the study.

Steps taken in phase 2 to maintain or increase reliability

I. INTRA-CENTRE RELIABILITY

Several means were used to maintain and possibly increase the reliability in ratings and diagnosis achieved in the preliminary phase of the study. One of the most important was that psychiatrists in the centres should carry out five simultaneous interviews at the beginning of phase 2 and continue to do a simultaneous interview each month during the course of the main phase of the project. Naturally during the course of the study some of the original raters left and new ones joined. Each new rater was trained in the centre, first by observing interviews carried out by those already familiar with the use of the PSE schedule and other instruments, and then by taking part in at least five simultaneous interviews with the original raters before starting to examine patients on his own. The extent of diagnostic reliability in the first 163 simultaneous interviews coming from these two sources is shown in Table 3.12. Roughly one in six of all interviews to date were simultaneous interviews, the others being consecutive interviews on the same patient or single interviews.

TABLE 3.12. *Agreement on diagnosis in simultaneous interviews in phase 2 of the IPSS*

Field research centre	*Number of simultaneous interviews*	*Agreement on 3-digit* ICD *categories* (*e.g. schizophrenia*)	*Agreement on 4-digit* ICD *categories* (*e.g. paranoid schizophrenia*)
Aarhus	11	9	7
Agra	23	23	18
Cali	11	10	7
Ibadan	19	18	11
London	14	13	10
Moscow	21*	20	16
Taipei	23	23	18
Washington	19	13	7
Prague	22	20	17
Total, all centres	163	149 94%	111 68·9%

* In addition to the above the Moscow Centre carried out 5 simultaneous interviews for which only one diagnostic assessment form was received.

The table shows the numbers of simultaneous interviews carried out in each centre. The lower numbers in Aarhus, Cali, and London indicate that the same raters were working throughout this phase. The over-all measure of agreement on the three-digit categories of the *ICD* was 94 per cent. Agreement on diagnosis according to the *ICD* classification was lower where a different classification had been in local use, although obvious errors in coding had been corrected at headquarters. This table corroborates the results of Table 3.11, that while a good measure of diagnostic agreement may be obtained on such broad categories as schizophrenia, manic depressive reaction, or reactive psychosis, there is less agreement on the finer categories in each group, represented by the four-digit classification, where the choice may depend on the relative importance attached by the rater to different symptoms.

2. INTER-CENTRE RELIABILITY

A number of exchanges of visits of collaborating investigators were carried out during phase 2 of the project and advantage was taken of these to test and improve the reliability of ratings made by psychiatrists from different centres. In some cases simultaneous interviews were done by two psychiatrists who came from different centres. Their ratings were discussed and causes of disagreement found. At other visits a number of psychiatrists from different centres simultaneously rated the same patient, using either live or video-taped interviews. Table 3.13 shows the results of the diagnostic evaluation of one of these video-taped interviews, made at the most recent exchange of visits, which took place at the end of field work in this phase.

The psychiatrists were first shown a video-taped recording of a PSE interview. After viewing the tape and rating the schedule, they were asked to make a diagnosis. Next, they were shown a video-taped psychiatric history interview of the same patient and asked to review their diagnoses. Finally they were given a narrative account of the patient's history as obtained from informants and hospital records, after which a final diagnosis was requested.

A few words should be said about the patient in question before discussing the results: this was a 41-year-old South African who had been living in the United Kingdom for about seven years. In the interview he described extremely vivid hallucinations in almost all sensory areas, connected with which he had a number of delusions. In the history interview he stated that he had been well until two years ago; he said he was married, well-off, had a doctor's degree in pharmacology, and no problems at work. The data from informants showed that he was in fact an undischarged bankrupt, separated from his wife, without a job, and without a degree in pharmacology; he even gave an incorrect family name on admission.

Table 3.13 shows that after the PSE interview all raters were agreed on the broad diagnostic group, in this case schizophrenia, and seven of the twelve agreed on the four-digit category of paranoid schizophrenia. After viewing the psychiatric history interview, these seven raters retained the same diagnosis, and were joined by three others. When presented with the informant's data, two of these three adhered to the diagnosis of paranoid schizophrenia and one failed to record an assessment. Only one psychiatrist of those who had previously diagnosed paranoid schizophrenia at both stages changed his diagnosis as a result of the informant's statements, although another rater

TABLE 3.13. *Diagnoses made by twelve psychiatrists simultaneously rating video-taped interviews, when provided with information from various sources*

Field research centre	*Psychiatrist*	*Diagnosis after viewing PSE interview only*	*Diagnosis after subsequently viewing psychiatric history interview*	*Final diagnosis based on information from PSE, PH, and informant*
Aarhus	012	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia
	013	295·9 Unspecified schizophrenia	781.8 Hallucinatory state	298.9 Reactive psychosis
Agra	021	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia	292.0 Organic psychosis (?GPI)
Cali	031	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia
	032	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia
Ibadan	042	295·9 Unspecified schizophrenia	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia
London	053	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia
Taipei	071	295·7 Schizo-affective sch.	295·3 Paranoid schizophrenia	—
	072	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia
	073	295·9 Unspecified schizophrenia	295·9 Unspecified schizophrenia	295·3 Paranoid schizophrenia
Washington	081	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia, 301.0 Hysteria
WHO, HQ	001	295·7 Schizo-affective sch.	295·3 Paranoid schizophrenia	295·3 Paranoid schizophrenia
Majority diagnosis		58% Paranoid schizophrenia	83% Paranoid schizophrenia	75% Paranoid schizophrenia

added hysteria as a concurrent condition. Six out of the twelve psychiatrists maintained the same diagnosis throughout the exercise. Dr Wing has recorded his impression that Maudsley-trained psychiatrists seem to use mainly PSE findings as the basis for their diagnoses, paying less attention to the patient's psychiatric history, but this may well be true of those trained in other schools.

Reliability in rating symptoms

It may be expected that psychiatrists will show better agreement on diagnoses than on their ratings of individual symptoms. For example, although two psychiatrists may differ in their descriptions of a particular type of delusion, they may nevertheless both take the fact that the patient was deluded into account in the same way in making their diagnostic assessments. Similarly, they may agree about the more extensive descriptions of behaviour, such as retardation or agitation, while disagreeing about smaller components of behaviour, such as long pauses between words or hand-wringing.

To test the reliability of ratings at the symptom level, the items in the PSE were formed into groups such that a positive reply on any of the items in the group could be taken to indicate the presence of a certain symptom, while a negative reply on every item would indicate its absence. At first the items were grouped purely on the basis of clinical judgement but, using data from the first 700 patients in the study, they were tested and regrouped several times, so as to achieve the best degree of association between the items in each unit. The groups so obtained, referred to here as units of analysis, may contain one or several items from the PSE schedule, items referring to observed behaviour being included.

The measure of disagreement used in this analysis was the 'mean serious disagreement index' (3). For this purpose a disagreement is regarded as serious if one rater says a symptom is absent and the other that it is present in some degree, but as not serious if they disagree about the severity or continuity of presence of the symptoms. The index is defined as the number of units with severe disagreement divided by the total number of units in which either one or both ratings are positive. It should be especially noted that in this index no attention is being paid to the number of units in which the psychiatrists agreed on the absence of symptoms, which in this study was very high. If the disagreement index were calculated on the total number of paired ratings including such double negatives, the indexes would appear very much lower than the ones shown in Table 3.14.

TABLE 3.14. *Index of mean 'serious' disagreement between raters in simultaneous interviews*

Field research centre	*Rater 1*	*Rater 2*	*Number of interviews*	**Mean serious disagreement index*
Aarhus	012	013	6	0·45
	013	016	3	0·47
Agra	021	022	12	0·28
Cali	031	032	8	0·47
Ibadan	042	041	5	0·39
	042	044	8	0·36
	042	049	4	0·35
London	051	053	9	0·31
Moscow	064	063	4	0·07
	064	065	2	0·17
	061	063	1	0·02
Taipei	071	072	9	0·12
	071	079	6	0·17
	072	079	4	0·21
Washington	082	083	11	0·53
	082	081	2	0·35
Prague	091	092	4	0·37
	092	093	3	0·48
	094	096	5	0·26
	094	097	4	0·14
†U.S./U.K.	A	B	20	0·32
Diagnostic	B	C	10	0·33
Project	A	C	7	0·48

* Mean Serious Disagreement Index $= \dfrac{\text{Disagreement Positive–Negative}}{\text{Total Units with Positive Ratings}}$.

† Items instead of Units of Analysis used.

The first line of the table shows that in the Aarhus centre the rater 012 did 6 simultaneous interviews with rater 013 and their 'mean serious disagreement index' was ·45. This indicates that for every 100 units in which at least one psychiatrist found the symptom present, the other disagreed with him in 45. In all, the ratings of 20 pairs of raters in 110 simultaneous interviews were analysed and the mean serious disagreement index varied from ·02 to ·53. This result compares reasonably well with the values obtained in the U.K./U.S. Diagnostic Project, where individual items from the PSE were used instead of the item-groups (units of analysis) used in the IPSS.

Bias of raters

In analysing the reasons for the disagreement between raters, special attention was paid to detecting the presence of bias in the interviews. One aspect of the raters' bias is a tendency to make consistently higher (or lower) ratings than the others in simultaneous or multiple interviews. Many explanations are offered for this type of bias: one is a difference in the training of psychiatrists. If training plays a role, one would expect all raters who belong to the same school of psychiatry to be consistently similar but possibly to differ from those who belong to other schools. Another reason for this type of bias can be the 'personal style' of a rater, in which case one would expect such a psychiatrist to be consistently different from others, regardless of type of training. For example, some psychiatrists are much more tolerant of abnormal behaviour than others and therefore give it consistently lower ratings.

To assess whether there was such a type of bias in any of the psychiatrists in this study, and whether it was present in one, some, or all areas of assessment, a comparison between their ratings in simultaneous and multiple interviews was made.

Comparisons between raters were made on the basis of pairing each psychiatrist with all the others with whom he participated in interviews. Thus if he took part in a simultaneous interview this counted as one comparison, but participation with four other psychiatrists in a multiple interview counted for him as four comparisons. Four groups of items were selected for these comparisons, relating to hallucinations (39 items), delusions (84), affect (36), and observed behaviour (105).

An index showing the tendency of a psychiatrist to rate higher or lower than his pair was calculated, based on the concept of positive and negative comparisons. For example, when the rater 012 was compared with 013 in assessing hallucinations in one of the interviews, it was found that 012 rated higher than 013 on 6 items and lower than 013 on 4 items, left 2 items blank, and rated 27 items equally with 013. Since the number of higher ratings exceeded the number of lower, this comparison counted as positive. The index of bias for any rater is then given by:

$$\frac{\text{No. of positive comparisons} - \text{no. of negative comparisons}}{\text{Total no. of comparisons}} \times 100.$$

From Table 3.15 it appears that in the four areas of psychopathology examined, 012 had a tendency to rate consistently

TABLE 3.15. *Bias of raters in inter-centre reliability interviews*

			Item groups			
Field research centre	*Rater*	*Number of comparisons*	*Hallucinations 39 items*	*Delusions 84 items*	*Affect 36 items*	*Observed behaviour 105 items*
Aarhus	012	57	+28	+19	+52	+19
	013	28	+50	+42	−14	−21
Agra	021	55	+10	+43	−21	+20
	022	55	−32	+52	+2	+18
Cali	031	58	−27	−34	−8	−34
	032	62	+14	+69	−35	−29
Ibadan	041	26	+23	−61	−57	−3
	042	55	−16	−10	0	+12
London	051	37	+5	−10	+24	+13
	052	38	−2	+39	+34	−3
Moscow	061	31	−48	−96	−3	+9
	062	31	−48	−12	−4	+17
Taipei	071	57	−31	−12	−3	+5
	072	54	−22	−38	−7	+51
Washington	081	26	−80	−7	+38	+30
	082	26	−30	−30	+38	−7
Prague	091	51	−3	+13	+19	−62
	092	31	+70	+25	0	−19
WHO, HQ	001	31	+25	+6	−38	−58
	002	39	−38	−74	−15	+33

higher than other raters in the study with whom he was compared whereas 031 rated consistently lower. It is of interest that both raters in the Moscow and Taipei centres tended to rate lower than others with whom they were compared on hallucinations, delusions, and affect and higher than others on observed behaviour; this may be accounted for perhaps by the raters in each centre belonging to the same school. Raters in the Agra centre also showed a tendency towards higher ratings in observed behaviour, whereas those in Cali and Prague tended to rate lower. It is possible that cross-cultural differences in tolerance of abnormal behaviour may have some effect in this area and when the whole material is available it will be possible to break it down to show the differences in bias when within-centre and between-centre comparisons are made.

TABLE 3.16. *Bias of raters in assessing delusions at four successive time periods. A: number of comparisons; B: index of bias*

		Timing and type of study							
		Phase 1				*Phase 2*			
Field research centres	*Rater*	*Within centres*		*Between centres*		*Within centres*		*Between centres*	
		A	*B*	*A*	*B*	*A*	*B*	*A*	*B*
Aarhus	012	11	+9	57	+19	6	+16	15	+20
	013	4	+25	28	+42	9	−22	15	−46
Agra	021	8	+75	55	+43	11	+27	15	+26
	022	8	−75	55	+52	11	−27	Not rated	
Cali	031	8	0	58	−34	8	−25	15	+13
	032	8	0	62	+69	8	+25	15	−66
Ibadan	041	8	0	26	−61	5	+40	Not rated	
	042	15	−60	55	−10	17	−17	Not rated	
London	051	5	+60	37	−10	9	+11	8	+62
Taipei	071	9	+33	57	−12	15	+26	15	+73
	072	9	−33	54	−38	13	+15	Not rated	
Washington	082	7	−50	26	−30	13	−30	15	−53
Prague	091	9	+22	51	+13	4	0	Not rated	
	092	9	−22	31	+25	7	−14	Not rated	

To find out whether raters have a consistent bias in rating a particular area of psychopathology, it is necessary to examine their ratings at different points in time.

Table 3.16 compares the bias of psychiatrists in rating delusions in groups in interviews done during the two phases of the study. It can be seen that several raters maintained a bias compared with pairs from both their own and other centres. For example, 012 and 021 were consistently higher raters throughout, whereas 082 was a persistently lower rater. Similar analyses were done for other groups of items and it was found that a persistent tendency to rate higher or lower, when present, seemed to be limited to certain areas of psychopathology only. Further analyses of both these types of bias will be carried out on the whole material, so that if necessary some form of correction of the ratings of the psychiatrists concerned may be undertaken.

Discussion

The results presented here give only an indication of what is being done in this study to test and improve reliability in rating psychiatric interviews. Methodology is being tested so that results based on the complete data of both phases of the IPSS may be more quickly obtained. Nothing has been said here about reliability testing of the psychiatric history and social description schedules, although some work has already been done on this very complex problem. Since information from these documents is used together with that from the PSE, in deciding upon the diagnosis, it is essential that they should satisfy the criteria for standardized research schedules already described in this paper and testing of their reliability must therefore be undertaken.

It is natural that a study like the IPSS should produce some side results. One of the more interesting of these is the effect it has had on psychiatrists using the PSE schedule. In the first place they found it was easy to adopt the procedure, because it was very similar to what would take place in a meticulously performed clinical interview. As time passed, they found themselves using PSE probes even when they did not have a schedule in hand and were not examining project patients. There arose a tendency to make their routine interviews resemble PSE interviews. In teaching undergraduates, they found themselves using PSE definitions or examples to explain phenomena of psychopathology. Such a use of PSE terms as a common frame of reference may make communication between psychiatrists with different backgrounds and mother tongues easier and more precise. The use of the same psychiatric examination schedule may, in a sense, provide a kind of common interviewing language, which will help to ensure that uniformity in case-finding which is an ultimate objective in epidemiological research.

REFERENCES

1. LIN, T. Y. (1967). 'The epidemiological study of mental disorders', *WHO Chronicle, 21*, **12**, 509–16.
2. WING, J. K., BIRLEY, J. L. T., COOPER, J. E., GRAHAM, P., and ISAACS, A. D. (1967). 'The reliability of a procedure for measuring and classifying present psychiatric state', *Brit. J. Psychiat.* **113**, 499–515.
3. KENDELL, R. E., EVERETT, B., COOPER, J. E., SARTORIUS, N., and DAVID, M. E. (1968). 'The reliability of the Present State Examination', *Social Psychiat.* **3**, 123–9.

Discussion of previous three papers

MARTIN M. KATZ
Chief, Clinical Research Branch, National Institute of Mental Health, Chevy Chase, Maryland, U.S.A.

Dr John Wing has reported on the development of a method for describing and classifying psychopathology, a relatively new technique which can be usefully applied to a wide range of problems in the fields of clinical research and epidemiology. Dr Cooper's paper describes its application in what can be properly called a cross-national investigation of the phenomena of psychopathology. Both the needs of epidemiology and the discipline of psychopathology are cast into somewhat bolder relief in this type of research. As had been indicated so often at this conference, both disciplines suffer from a lack of reliable methods for the objective description and quantification of the various facets and classes of disordered behaviour.

I will make a few comments on methodologic issues presented by this new technique and then turn to a consideration of the results of Dr Cooper's study.

There are, as many of you are aware, a very large number of observational rating methods for describing and quantifying psychiatric symptomatology. They vary widely in applicability and in quality. The most popular of these have been developed in accord with sound psychometric principle and have been exceedingly useful in research, particularly as regards the problem of evaluating psychiatric and psychological treatments. Lorr's In-patient Multidimensional psychiatric rating scale (5) and Wittenborn's psychiatric rating scale (7) are examples of the better techniques in this area. Useful as these instruments are, Dr Wing rightly points out that they do not provide a systematic or standard protocol for the collection of data on patients as part of their administration and are therefore less useful in situations where the *type* of interview takes on great importance, e.g., in the cross-national study, where the research interview in the different settings can be expected to vary markedly in the kind of information which is sought and in the style in which it is carried out.

The development of a 'standard interview' was, in fact, the next natural or logical step to take in the movement to bring a more systematic, quantitative approach to the problem of describing and classifying psychopathology. It was, however, not a very easy one to accomplish. Both Dr Spitzer (6) and Dr Wing can testify to the arduous nature of the task of developing a 'standard' script to be followed in the cross-examination of a patient and the complications they faced in trying to effect a workable compromise, i.e. an interview that would meet the necessary research requirements yet one that the practising clinician would not feel does violence to the clinical process. The two sets of aims represented by the clinical and the research points of view are, as we all know, not necessarily compatible. Rather than satisfy either camp, you are likely to have to put up with a good deal of criticism from both for either being too rigid and thus stultifying the clinical process, or for being too loose and thus permitting too much variability in the approach. Nevertheless, speaking as a researcher, I think that a standard interview technique is essential for certain kinds of research and that its development will contribute substantially to reducing error and increasing sensitivity in the experimental attack on a whole range of clinical research problems.

Given the wide applicability of the psychiatric rating method, I find myself differing from Dr Wing when he discusses the 'dimensional' versus the 'medical' model and the manner in which he expects his method to contribute to problems in diagnosis.[1] I am not clear, for example, on what he has against the concept of developing reliable measures of 'dimensions' of psychopathology. There may be some semantic difficulty that I am tripping over here, but the need for the development of empirical measures of *facets* of behaviour is as pressing as the need to develop a new method for diagnosing or classifying patients. In fact, the strength of an empirical technique of this type lies in its capacity to slow down the process of achieving what appears to be premature closure of the establishment of a system of classification for the behaviour disorders. Such techniques are likely to make us more, rather than less, aware of which facets of abnormal behaviour are poorly understood, inadequately articulated, and thus unlikely to be judged reliably by clinicians. These 'unreliable' facets can turn out to be very critical factors in differential diagnosis, e.g., apathy, as we have found in some previous research (3). The use of such techniques permits this type of detection and thus opens the way for more focused, intensive study of these factors and the process of diagnosis.

Dr Wing raised a concern that if a new classification system which results from such an empirical approach is not sufficiently like the traditional one, how will we assess its validity? I agree that is a major problem, but I think a soluble one. Whether it is soluble, however,

---

[1] I have discussed the role of the 'dimensional' model in more detail in a different context in a recent article.

should not really be an issue. The interview method is designed to result in some new data and new ideas on classification. I think there is sufficient evidence of a negative nature about the current nomenclature to warn us to carry the traditional classification system 'lightly', and to take advantage of the potential of this basically empirical method for enlightening us further about psychopathology —in other words, to permit the new method to lead rather than to follow in this area.

Dr Cooper and his colleagues, Dr Zubin and Dr Gurland, apply the technique to a very interesting research problem in diagnosis and psychopathology, i.e., the sources of variation in American and British mental hospital statistics of the number of admissions for schizophrenia and the affective disorders. Do Americans have proportionately more schizophrenics than the British? Do the British have significantly more affective disorders than the Americans? The answer in both cases is 'not very likely'. It is probable that such differences in admission statistics can be traced to variations in diagnostic practices in the two countries (one of the three possible sources of variance that Dr Cooper mentions) and not to real differences in the patients themselves. That American and British patients are qualitatively different seems very unlikely. I would, however, ask you to keep your mind open on such questions. As documented by serious students of the British and American scenes, e.g., Gorer (1), the British people and the American people are not all that alike: there is a distinct possibility that they could produce different kinds of patients.

The question of differences in 'diagnostic practices' also has two facets which must be considered. The differences could, for example, be semantic, i.e., problems in definition, differences in how these disorders are defined in the two national settings. Better communication between clinicians or more precision in terminology should eventually resolve this kind of problem. On the other hand, it may be that British and American psychiatrists actually perceive patients differently and judge the quality of emotions differently, i.e., that we are dealing with a national or ethnic difference in the perception of behaviour and emotion. That may sound an unlikely possibility—but it is important to keep in mind that it could be so. We had an example of this several years ago when we asked large samples of American and then British clinicians to view the same patient interview (a film of a clinical interview with an American patient) and both to provide a diagnosis and to rate the patient's symptoms and behaviour on a standard psychiatric rating scale (3). To no one's great surprise, the distributions of diagnoses in the two groups were very different. One-third of the forty Americans had called the patient 'schizophrenic'—

not one of thirty-two British psychiatrists had. The majority of the British had given the patient an 'affective' diagnosis, e.g., 'emotionally unstable' or 'neurotic depression'. Equally important, when judgements of specific symptoms were analysed it was found that the Americans, as one might expect, saw significantly more pathology in the patient than did the British. The latter exceeded the Americans in one area only and that was on a factor which primarily measures depression. If one adjusts the individual factor scores for the difference between the British and the Americans in total amount of psychopathology seen, then this difference between the two in the amount of 'depression' judged to be present in the patient, takes on much greater significance in the comparison. The result indicates that the British saw a good deal more depression in the patient than did the Americans, a finding which appears to be compatible with the British tendency to diagnose proportionately more affective disorders than the Americans.

This was only one case. The main result, has however, something in common with the hospital admission statistics and is presented here only to open minds to the possibility of the existence of differences in the ways in which clinicians from different national backgrounds may perceive patient behaviour and emotion. Since then, Drs Cooper, Gurland, and Sharpe have carried out more studies on this problem and are prepared to offer more comprehensive information in this area.

The clinicians themselves are, therefore, obviously one source for the variation in hospital statistics on diagnosis between the countries. Another probable source, as we have pointed out, is in the patients. It may be that British schizophrenic and depressed patients actually manifest their psychoses differently from American patients. As I noted, this result is not so unlikely either. I would like to present a case in point why, despite the similarity of these two countries and their people, this possibility needs to be taken seriously. We have a situation in another study where two ethnic groups from the same community (Hawaii), who enter the same mental hospital, are being studied as regards their patterns of symptoms and behaviour (4). In this case they come from the same community, enter the same hospital and are interviewed by the same psychiatrists—in a situation where the nature of the interview is controlled (we used the Spitzer Mental Status Interview) and in which standard psychiatric ratings are used. The two groups are Hawaii/Japanese and Hawaii/Caucasian and, as you will see in Fig. 3.1, they are strikingly different. The Caucasian psychosis is more affective in quality and more hostile and depressed; the Japanese is more emotionally withdrawn, more retarded, and more disturbed in basic thought processes. These are

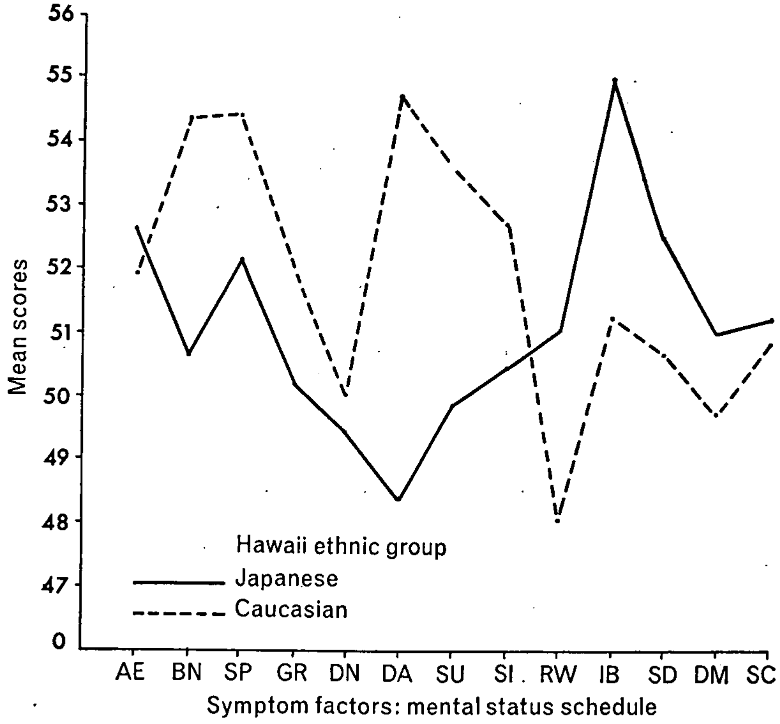


FIG. 3.1. Hawaii/Japanese and Hawaii/Caucasian functional disorders: patterns of symptomatology in the hospital.

Note: AE = agitation-excitement; BN = belligerence-negativism; SP = suspicion-persecution-hallucinations; GR = grandiosity; DN = denial of illness; DA = depression-anxiety; SU = suicide-self mutilation; SI = social isolation; RW = retardation-emotional withdrawal; IB = inappropriate-bizarre; SD = speech disorganization; DM = disorientation-memory. (Reproduced from Katz, Sanborn, and Gudeman, 1969.)

patients with different ethnic backgrounds who reside in the same community. There is some difference in the patterns of diagnoses for these two groups but the major source of the difference lies in the attributes of the patients themselves. In this study our major objective is to characterize the psychopathology of the respective ethnic groups. The fact that specific ratings of symptoms and behaviour are available permits the diagnostic issue to be set aside and the central aim of the study to be approached directly. How does all this relate to Dr Cooper's results and conclusions?

(1) When the standard interview procedure is conducted in both the Brooklyn and the Netherne settings, the differences in diagnoses are reduced markedly but they do not disappear. Proportionately

more depressions are still seen at Netherne and the difference with regard to schizophrenia, although markedly reduced, borders on significance.

(2) That this result is not peculiar to the Netherne/Brooklyn comparison is clear when one compares statistics from representative samples of hospitals in London and New York; the differences in the frequency of diagnosing 'depressions' still persist.

Dr Cooper, Dr Zubin, and their colleagues have apparently controlled for variations in the diagnostic interview, which leaves, as possible sources of explanation, variations in the perception of the respective clinician groups or real differences in the patients, i.e., Britain simply has more depressives entering hospitals than has the U.S. I would suggest that the next step here is to set aside the diagnostic analysis and to take advantage of the method and look directly at the specifics and the patterns of symptoms and behaviour in the two national samples.

On this point I had difficulty in interpreting the results with the computer diagnosis 'Diagno'. My interpretation of these results is that the computer program, on the basis of the same data that the clinicians were using to arrive at a diagnosis, simply demonstrated that it would have diagnosed many more schizophrenics and many fewer depressions both in Britain *and* in the U.S. than did the project psychiatrists or the U.K. hospital psychiatrists.

In the case of the psychiatrists, it indicates that despite the fact that they were apparently working with the same data that the computer was confined to, their intuitive processes were probably using different criteria to arrive at a diagnosis. There are apparently other influences at work which account for these contradictions, and a further, more intensive, look at possible sources may be very useful in understanding the limitations on both sides. I do not think that the analysis of the ratings presented by the authors does very much to clear up these contradictions. It does, however, make the direct comparisons of the symptom patterns of the two national groups (the diagnostic issue aside) all the more interesting to look forward to.

All in all, I am very reassured, as I think we all should be, that with new methods of this type we can begin to study the diagnostic process more intensively and can begin to separate out and assess the various elements which significantly influence the results of cross-national and cross-ethnic studies. Real progress is thus being accomplished toward developing a clearer picture of the phenomena of psychopathology.

REFERENCES

1. Gorer, G. (1964). *American People—A Study in National Character* (New York: W. W. Norton).
2. Katz, M. M. (1968). 'The dimensional and typological approaches to assessment and prediction in psychopharmacology', *Psychopharmacology Review of Progress 1957–67* (GPO-PHS Publication no. 1836).
3. —— Cole, J. O., and Lowery, H. A. (1969). 'Studies of the diagnostic process: the influence of symptom perception, past experience and ethnic background on diagnostic decisions', *Amer. J. Psychiat.* **125**, 7.
4. —— Sanborn, K. O., and Gudeman, H. (1969). 'Characterising differences in psychopathology among ethnic groups in Hawaii', *Social Psychiatry* (from the 47th annual meeting of the ARND (Baltimore: Williams and Wilkins).
5. Lorr, M., McNair, D. M., Klett, C. J., and Lasky, J. J. (1962). 'Evidence of ten psychotic syndromes', *J. consult. Psychol.* **26**, 185–9.
6. Spitzer, R. L., Fleiss, J. L., Burdock, E. I., and Hardesty, A. S. (1964). 'The Mental Status Schedule: rationale, reliability and validity', *Compr. Psychiat.* **5**, 384–95.
7. Wittenborn, J. R. (1955). *Manual: Wittenborn Psychiatric Rating Scales* (New York: Psychological Corporation).

4

PSYCHIATRIC DISORDERS IN CONTRASTING CLASS AND ETHNIC GROUPS

Measures of psychiatric disorder in contrasting class and ethnic groups

A preliminary report of on-going research

BRUCE P. DOHRENWEND
Associate Professor
EDWIN T. CHIN-SHONG
Research Associate
GLADYS EGRI
Instructor
FREDERICK S. MENDELSOHN
Research Associate
JANET STOKES
Assistant

Department of Psychiatry, College of Physicians and Surgeons, Columbia University, U.S.A.

The methodological study described in this report focuses on the central unsolved problem of psychiatric epidemiology—the problem of how to conceptualize and measure psychiatric disorder independently of treatment status. The aim of the study is to develop better measures of psychiatric symptomatology and related disability in role functioning than have been used heretofore in epidemiological studies of general populations.

The long-range goal of the research program of which this study is a part, however, is to bring these measures to bear on a major substantive issue: the problem of the relative importance of social causation *v.* social selection factors in social class differences in rates of psychopathology. Our formulation of the theoretical assumptions and research strategy for the investigation of this issue has been set forth in detail elsewhere (4).

The problem of assessing psychiatric disorder in general populations

Most studies of relations between social factors and psychiatric disorders have defined disorder in terms of admission to treatment. While operationally clear, such a definition of a 'case' is also one of the most limited. Its limitations are especially evident when research focuses on the possible significance of social factors in aetiology. Treatment rates vary, for example, with the availability of facilities and with public attitudes toward their use. Either could be responsible for spurious relations between social factors and rates of disorder measured by number of cases in treatment (4, chapter 1).

Recognition of such problems has led at least thirty-five different investigators or teams of investigators to attempt to count untreated as well as treated cases of psychiatric disorder in forty-four community studies (4, chapters 2 and 7). With very few exceptions, cases in these studies have been defined by applying clinical judgements to symptom data.

The investigators who have conducted these studies have asked that their determinations be taken largely on the basis of the clinical training and experience of the participating psychiatrists. This is true, by and large, even of two most methodologically explicit and ambitious of these works—the Midtown (17) and Stirling County (13) studies. With a few marked exceptions such as the consistent finding of the highest rate of disorder in the lowest social stratum, however, these studies present a picture of inconsistent results (4, chapters 2 and 7). And since the evidence for the validity of the measures of disorder used is sparse in all of these studies, it is impossible to choose the more valid among them as a source of measures for further research (4, chapter 7).

One way to throw light on this methodological problem is to compare and contrast different types of measures of psychiatric disorder under controlled conditions. An ideal procedure would be to test the different measures against an ultimate criterion of the presence or absence of psychological disorder in cross-sections of general populations. Failing this, a decisive procedure would be to test the ability of the different measures to discriminate between a known 'ill' group of patients, representing the full range and variety of disorder on the one hand, and a known 'well' group, representing on the other hand the full range and variety of 'wellness'. Neither of these procedures is, of course, possible to implement. First, there is no infallible test of the presence of those types of psychological disorder that are thought to be functional—the types of greatest interest in the present focus. Second, there are no criteria for pre-selecting a control group that represents all varieties of 'wellness'. Finally, one of the questions at issue is precisely to what extent patient criterion groups do represent the range and variety of disorder existing in the population. How are we to come to terms with these rather harsh realities?

Method of procedure

The procedures used for measuring psychiatric disorder in previous studies have been of three main kinds. First, as in most of the European and Asian research, a single psychiatrist (or sometimes a small team of psychiatrists) has collected data on community populations and has made psychiatric assessments of these data. As a rule, however, the data collection procedures have not been made explicit in this type of study. Second, as in the Stirling County and Midtown investigations, standard and explicit data collection procedures have been used, but the interviewers have not been psychiatrists or clinical psychologists; rather, the psychiatrists on the staffs of these studies have made assessments of the written records of the data collected by lay interviewers.[1] And third, an objective measure has been used (14) which consists of a 22-item screening instrument that was developed on the basis of the high correlations of those items with the judgemental ratings made by psychiatrists on the Midtown study (12). In almost all of the previous studies the problem of testing the criterion-oriented validity of the measures

---

[1] An exception is a study of the Yoruba by A. H. Leighton and his colleagues in which written records of interviews collected by standard procedures were evaluated by several psychiatrists, but with the initial interviews conducted by psychiatrists. See A. H. Leighton, T. A. Lambo, C. C. Hughes, D. C. Leighton, J. M. Murphy, and D. B. Macklin, *Psychiatric Disorder Among the Yoruba* (Ithaca, N.Y.: Cornell University Press, 1963).

was either ignored (the rule), or focused on small, diagnostically homogeneous groups of patients.[1]

In the present study, conducted in the Washington Heights district of New York City, our choice of subjects represents an attempt to provide sharp and clear contrasts that cannot help but be informative. Thus, for example, a number of community leaders—undoubtedly far from 'symptom-free', but shown by their reputation and performance to be highly effective citizens—provide contrast with heterogeneous groups of psychiatric patients. And instead of the yet-to-be-developed 'infallible' test for psychological disorder, or the impossible-to-use clinical ideal of diagnosis over a course of therapy, we rely again on sharp contrasts, this time by experimentally comparing two different data collection instruments. One is a Midtown and Stirling type of interview that we call the Structured Interview Schedule; the other is the Psychiatric Status Schedule developed by psychiatrists and psychologists of Biometrics Research at the New York State Psychiatric Institute (16). Moreover, unlike the Midtown and Stirling County studies, but consistent especially with most of the European and Asian studies, the present research has psychiatrists conducting the interviews. Let us describe these procedures in more detail.

THE SUBJECTS

The subjects in the present study are between the ages of 21 and 64 and come from five ethnic groups: white Anglo-Saxon Protestants or close approximations thereof, Jews, Irish, Negroes, and Puerto Ricans. Our aim was to interview a total of about 580 individuals, of the following composition: 240 psychiatric patients, chiefly out-patients but also some in-patients and some convicts; 60 community leaders; and 280 present or recent residents of Washington Heights (only a small minority of whom are or have been psychiatric patients), sampled on a probability basis from each of the five ethnic groups.

---

[1] A possible partial exception is the study by Manis *et al.* (14), where representative samples of about 100 patients each were drawn from the receiving ward and the predischarge ward respectively of a state mental hospital. However, the diagnostic composition of these patients was not reported. See Manis *et al.*, 'Validating a Mental Health Scale', *American Sociological Review*, **28** (1963), 108–16.

To achieve this goal of obtaining about 580 interviews, we allowed for refusals and other sources of non-response by selecting more potential respondents than we would be able to successfully interview. We expected losses of about 25 per cent to 30 per cent on the basis of past experience (3, p. 26; 7, p. 89).

TABLE 4.1. *Outcome of interview attempts according to type of respondent (percentage)*

	Type of respondent		
Outcome of interview attempt	*Leaders*	*Ethnic samples from community*	*Patients (including out-patients, in-patients, and prisoners)*
Completed interview	78·8	65·5	75·7
Partial interview	3·5*	0·3	3·8†
Refused by respondent‡	8·2	23·0	7·9
Refused by other on behalf of respondent	1·2	1·5	0·8
Other losses:			
Death	3·5	2·3	—
Serious illness or injury prohibiting interview	1·2	0·5	—
Verified move out of region	1·2	4·9	2·1
Unable to locate	2·4	2·1	9·2
Total %	100·0	100·1	99·5
Total respondents from whom interviews were sought	(85)	(391)	(239)

* These were broken off after the first few questions and provide too little information for useful analysis.

† These are in-patients who were too disturbed to be interviewed in the usual way. Eight of the nine patients in this category gave information that could be used for construction of some objective measures.

‡ Including those not obtained because of numerous broken appointments.

As Table 4.1 shows, we did slightly less well among the community respondents than we expected even though, since these were a probability sample, we tried much harder to convert them than we did for the patient sample and even for the leaders.[1] Preliminary analyses of previous data on non-respondents among the community ethnic samples and the patient groups are under way. Thus far, there seems to be little difference between those obtained and those not obtained on the kinds of symptoms (see the first 22 items listed in Table 4.8) that we have examined so far.

Patients

The 'patients' come from three main types of facilities: psychiatric out-patient clinics, mental hospital in-patient services, and prisons.[2] The large majority, however, are from the first of these sources, as Table 4.2 shows, and were interviewed in their homes as were the ethnic samples from the community and the community leaders.

TABLE 4.2. *Different sources from which patients were drawn according to their ethnic backgrounds (percentage)*

	Ethnic background					
Source	*White Protestants of old American ancestry*	*Jewish*	*Irish*	*Negro*	*Puerto Ricans*	*Total*
Out-patient clinics	47·1	64·7	50·0	55·0	71·9	58·9
In-patient facilities	29·4	27·5	30·0	25·0	15·6	25·3
Alcoholic clinic	—	2·0	13·3	1·7	—	3·2
Prisons	23·5	5·9	6·7	18·3	12·5	12·6
Total %	100·0	100·1	100·0	100·0	100·0	100·0
Total respondents interviewed	(17)	(51)	(30)	(60)	(32)	(190)

As noted in Table 4.1, the problem of incomplete interviews was greatest in the in-patient facilities. Here, we sought to interview admissions that occurred within the previous week in

---

[1] Less effort at follow-up was made in the case of patients since they were not sampled on a probability basis and losses could, therefore, be replaced if necessary without compromising the logic of our procedures for selecting them.

[2] The out-patient services are those at Vanderbilt Clinic at the Columbia Presbyterian Medical Center, Harlem Hospital, and Jewish Memorial Hospital. The inpatient services are at Bronx State Hospital and Manhattan State Hospital. The prisons are Rikers Island and Women's House of Detention.

order to get more accurate recall of role behaviour in the community; in doing so, our psychiatrists found that some of the designated patients were too acutely disturbed to undergo the interview.

The patients are selected with the aim of an even distribution of six 'behaviour types' within each of the five ethnic groups. These 'behaviour types' are similar to the fictitious illustrations of different types of disorder developed by Star (19) and used in our own studies of public attitudes towards mental illness (5). The procedure has been to have a patient's therapist or evaluator (if the patient is on a waiting-list) rank-order, from most to least like the patient, cards containing these six descriptions abbreviated as follows:

(*a*) A person who is very suspicious; does not trust anybody; and is sure that everybody is against this person.

(*b*) A person who is very quiet, does not talk much to anyone; acts afraid of people; stays alone and daydreams all the time; and shows no interest in anything or anybody.

(*c*) A person who worries a lot about little things; seems to be moody and unhappy all the time; and cannot sleep at night, brooding about the past, and worrying about things that might go wrong.

(*d*) A person who drinks too much; goes on a spree when there is money in the pocket. This person promises to stop drinking, but always goes off again.

(*e*) A person who just cannot leave the house without going back to see whether the gas stove was left lit or not; always goes back again just to make sure the door is locked; and is afraid to ride up and down in elevators.

(*f*) A person who has been telling lies for a long time now, and stealing things. Others are very upset about these acts, but the person pays no attention to others.

The patient has then been selected (or passed by if enough had already been secured in a particular category) in terms of the card most like him or her. This procedure was designed to ensure a wide variety of types of disorder among interviewed patients, many of whom were on waiting-lists and for whom, therefore, there was not available a firm diagnosis.[1]

As Table 4.3 shows, white Protestant patients of old American ancestry proved scarce. There is also a tendency for the first three behaviour types to be more numerous than the last three. The reason for this is mainly that the first three ('suspicious', etc., and 'quiet, afraid', and 'worry, moody, anxious') were more likely to characterize hospitalized patients than the last three, and so were added from both out-patients and in-patients.

TABLE 4.3. *Selection of behaviour types according to ethnic backgrounds of patients (percentage; out-patients, in-patients, and prisoners combined)*

	Ethnic background					
Behaviour type	*White Protestants of old American ancestry*	*Jewish*	*Irish*	*Negro*	*Puerto Rican*	*Total*
Suspicious, etc.	23·5	28·8	24·1	25·0	15·6	24·2
Quiet, afraid, etc.	29·4	23·1	20·7	16·7	28·1	22·1
Worry, moody, unhappy, etc.	11·8	21·2	17·2	18·3	25·0	19·5
Drinks too much, etc.	11·8	3·8	27·6	10·0	3·1	10·0
Checks stove, afraid of elevator, etc.	—	17·3	3·4	10·0	15·6	11·0
Lies, steals, etc.	23·5	5·8	6·9	20·0	12·5	13·2
Total %	100·0	100·0	99·9	100·0	99·9	100·0
Total respondents interviewed	(17)	(52)	(29)	(60)	(32)	(190)

Community leaders

In 1960, we completed interviews about attitudes toward mental illness with 87 community leaders in Washington

---

[1] Psychiatrists on the study and others consulted have agreed that the descriptions of the disorders described as functional in the *Diagnostic and Statistical Manual* of the American Psychiatric Association can be coordinated with these descriptions in terms of the cards they are most like. It will later be possible to check this out empirically by investigating the relation of the card sorts to firm diagnoses based on course of treatment.

Heights. These were selected on the basis of formal position of influence and on the basis of reputation among other leaders and all but a few were Jewish, Irish, Negro, or Puerto Rican. These leaders included state senators, assemblymen, municipal court justices, businessmen, school principals, clergymen, and heads of such organizations as the Chamber of Commerce, League of Women Voters, and Puerto Rican hometown clubs.[1]

Our present sample includes the top ten or so of the above leaders from each of the four ethnic groups as indicated by the number of nominations they received from other leaders and as cross-checked by nominations from a small probability sample of household heads interviewed in 1962–3.[2] To this group, we added additional leaders on the basis of new office-holders in positions that had yielded top leaders in 1960.

TABLE 4.4. *Types of leadership position by ethnic backgrounds of leader (percentage)*

	Ethnic background					
Type of leadership position	*White Protestants of old American ancestry*	*Jewish*	*Irish*	*Negro*	*Puerto Rican*	*Total*
Political–legal	—	17·6	66·7	37·5	15·8	26·9
Economic	33·3	11·8	—	12·5	—	9·0
Educational	16·7	29·4	—	—	—	9·0
Religious	50·0	17·6	22·2	18·8	15·8	20·9
Civic–social	—	23·5	11·1	31·3	68·4	34·3
Total %	100·0	99·9	100·0	100·1	100·0	100·1
Total leaders interviewed	(6)	(17)	(9)	(16)	(19)	(67)

In addition, several events that occurred in Washington Heights made it possible to extend coverage still further on the

---

[1] For a more detailed description of these leaders and our procedures for selecting them, see B. P. Dohrenwend, 'Some aspects of the appraisal of abnormal behavior by leaders in an urban area', *American Psychologist*, **17**, 190–8.

[2] This cross-check by a small sample of 151 community respondents added no new leaders; that is, no new persons emerged who had been missed by the nominations from leaders but who received nominations from more than one or two of the sample of 151 community respondents.

basis of the emergence of new Negro and Puerto Rican leaders along with new leaders from other groups. With the growth of the movement for civil rights, issues such as the 'bussing' of schoolchildren became focuses of controversy in Washington Heights. It was possible in the spring of 1966, by interviewing a number of the people most directly involved, to secure nominations of individuals who emerged as most influential on the various sides of these controversies and to add a 'civic–social' category of leaders on this basis.

Table 4.4 reflects the fact that in Washington Heights, only the leaders of Jewish background have had substantial positions and reputations for leadership in all five of the areas of leadership covered.

Of the 67 leaders with whom interviews were completed, the large majority had never been in treatment. Nevertheless, 14·9 per cent reported in response to direct questioning that they were or had been in treatment with members of the mental health professions.

Community samples

The probability samples of community respondents were stratified on educational level, an indicator of class,[1] in a way that sought to balance the unequal class distributions in each of five ethnic groups: white Anglo-Saxon Protestants or close approximations thereof, Jews, Irish, Negroes, and Puerto Ricans.

The stratified sampling plan for the ethnic samples of community respondents was possible to entertain because of the unusual research facilities of the Columbia Presbyterian Medical Center—more specifically, the Community Population Laboratory. This laboratory periodically conducts large-scale sample surveys. The first, in 1960–1, consisted of 4,500 households; the second, in 1965–6, consisted of 3,000 households (7).

---

[1] No attempt was made to combine the three most commonly used indicators of social class—education, occupation, and income—into a composite index because education has different implications for income according to ethnicity (see Dohrenwend and Dohrenwend (4), pp. 60–1). Educational level was used (rather than income or occupation) to stratify respondents *within* ethnic groups in selecting the samples because education is usually antecedent to occupation and income and is a standard with more in common for both men and women.

It is from these that we have drawn our samples of community respondents.

About half the subjects have been respondents in our previous research which began for some as early as 1960 (4, 5). On a small subsample of 41 respondents, indeed, we have data on a number of symptom items at three points in time between 1960 and the end of 1967.

Table 4.5 shows that despite our efforts, we are short of lower educated respondents among white Protestants of old American ancestry and to a lesser extent, among Jews. We are also somewhat short of college graduates among Puerto Ricans and among women who head households in which no husband is present. The main reason for these shortages is that these were the rarest categories in the population of Washington Heights.

Of the 257 community sample respondents with whom we completed interviews, 9·7 per cent reported in response to direct questioning that they were or had been in treatment with members of the mental health professions.

THE TWO INTERVIEW INSTRUMENTS

The two interview instruments, mentioned earlier, are called (for purposes of identification only) the Structured Interview Schedule and the Psychiatric Status Schedule. They are used with equivalent samples of subjects in order to elicit evidence of psychiatric symptomatology, and attendant impairment of functioning in work, marital and sexual relations, child-rearing, housekeeping, friendship, and leisure activities. Although they attempt to cover essentially the same areas, they contrast sharply in the way each goes about it.

Structured Interview Schedule (SIS)

The Structured Interview Schedule or SIS has been modelled by us after the types of questionnaires that were used in the Midtown and Stirling County studies, and in our own previous work in Washington Heights. Our reason for choosing the Midtown and Stirling County types of questionnaire as models for one of our alternative instruments is that, as was mentioned earlier, these two are among the very few field studies that provide a standard, explicit set of data for psychiatric assessment. Thus, the SIS is a conventional type of survey questionnaire, relying

TABLE 4.5. *Educational level of household head for respondents in the community ethnic samples (percentage; married females classified in terms of educational level of spouses)*

	Ethnic samples						
Years of formal education	*White Protestants of old American ancestry—male household heads and spouses if married*	*Jewish male household heads and spouses if married*	*Irish male household heads and spouses if married*	*Negro male household heads and spouses if married*	*Puerto Rican male household heads and spouses if married*	*Negro female heads of household with no husband present*	*Other female heads of household with no husband present*
0–7	7·7	4·8	18·8	40·0	15·4	29·4	24·3
8–11	3·8	28·6	27·1	20·0	41·0	20·6	37·8
12–15	53·8	40·5	31·3	20·0	33·3	38·2	24·3
16 or more	34·6	26·2	22·9	20·0	10·3	11·8	13·5
Total %	99·9	100·1	100·1	100·0	100·0	100·0	99·9
Total subjects interviewed	(26)	(42)	(48)	(30)	(39)	(34)	(37)

heavily on items with fixed alternative-response categories, e.g., 'true–false', drawn, as in the earlier studies, principally from the World War II Neuropsychiatric Screening Adjunct (18) (Stirling and Midtown), and the MMPI (2) (Midtown). The SIS is, however, much expanded over the previous versions, especially in its coverage of impairment of function in role areas.

The Psychiatric Status Schedule (PSS)

A distinctive feature of the SIS is that the response alternatives are almost always presented to the subject, who then chooses among them. The Psychiatric Status Schedule, or PSS, by contrast, relies mainly on open-ended questions, the probed responses to which are coded into fixed categories, descriptive of pathology, on the basis of the clinical judgement of the interviewer. Central to the SIS is a core of symptom items that have been shown in past research to discriminate on an actuarial basis between patient and non-patient groups.[1] The PSS, on the other hand, has its origin in the mental-status interview that is traditionally administered as part of clinical examinations. It is thus probably more like the clinical interviews that have been used in some of the previous field studies conducted by individual psychiatrists, but with an added attempt to standardize questioning and recording procedures.

SYSTEMATIC ALTERNATION OF THE SIS AND PSS AMONG RESPONDENTS

Because of the length of the interview which took, on the average almost an hour and three-quarters, and because of possible problems of order effects, only one of the two interview instruments was used with each respondent. Systematic alter-

---

[1] Our previous structured schedule, which was administered by lay interviewers to the more than 100 Jewish, Irish, Negro, and Puerto Rican out-patients and the 151 cross-section respondents, contained a number of symptom items, in addition to the 22-item Midtown screening instrument. (For a description of some of these, see Dohrenwend and Dohrenwend (4), chapter 6.) Analysis showed that 14 of these additional symptom items discriminated at the 0·01 level or better between patients and non-patients. These, as well as the 22 Midtown items, themselves chosen on an actuarial basis, are retained in the Structured Interview Schedule. In addition, some items from the Stirling County study and some items to elicit symptoms of certain kinds of psychotic behaviour were added, as well as the items on role-functioning mentioned above.

nation of the two schedules throughout each interviewer's assignment was designed, however, to give all subjects an equal chance to be interviewed with either one schedule or the other, thus permitting controlled comparisons of the results achieved with each schedule.

A 'SELF-ADMINISTERED' SECTION FOR BOTH THE SIS AND PSS

Social desirability ratings (6), collected from small samples of Jewish, Irish, Negro, and Puerto Rican community members, have been analysed and the results used to match items (e.g., on fear, different modes of expressing anger, and different 'behaviour types', on levels of social desirability). A number of 'socially undesirable' items have been paired in a forced-choice format, together with a third neutral or 'desirable' decoy. These have been self-administered as part of the present interviews, so that it is possible to look at responses to these items that are either controlled in this manner for social desirability, or uncontrolled, as they were presented in the previous lay interview, where the items were administered in a true–false format, with alternated keying to test for acquiescence. Our inclusion of the triads of forced-choice items reflects our continuing concern with the problem of response bias and the need to investigate the part it may play, especially in answers to the SIS. A large number of the closed fixed alternative items in the SIS may be maximally susceptible to both acquiescence and social desirability response styles.

A 'STRESSOR INVENTORY' SECTION FOR BOTH THE SIS AND THE PSS

At the end of both schedules, there is the following question: 'What was the last major event in your life that, for either better or worse, interrupted or changed your usual activities?' and the probe: 'For example, events affecting your occupation, your physical health, your living arrangements, your relations with other family members, your friends or your personal values or beliefs.' This is followed by further probes for details, and an open-ended question about other events almost as important since then. In addition, a checklist of events that may have occurred during the last 12 months is administered, along with a checklist of events that the respondent anticipates will occur during the next 12 months. Examples of such events are 'married', 'birth of first child', 'death of a loved one', 'fired', and so on. These questions constitute a preliminary attempt to develop a stressor inventory, and reflect our concern with the bearing of situational factors on psychological symptomatology (4, chapters 8 and 11).

THE INTERVIEWERS: PSYCHIATRISTS *v.* LAY INTERVIEWERS

The interviews with the above instruments have been conducted by fifteen psychiatrists, all but one of whom had completed their residency training. They received intensive instruction averaging 30 hours in the use of these instruments. Initial interview assignments were randomized.

RECORDING OF INTERVIEWS

All interviews were recorded on interview forms in which most categories of response or judgement are precoded. In addition, attempts were made to tape-record all interviews as a matter of routine unless the respondent refused permission or unless the psychiatrist thought either from the beginning or during the interview that reactions to the recording would harm the quality of the data. The large majority, 84 per cent of the interviews, were successfully tape-recorded.

DEVELOPMENT OF JUDGEMENTAL AND OBJECTIVE MEASURES OF PSYCHIATRIC DISORDER

From the data secured with the various instruments, a number of different measures are being developed. These include judgemental ratings made according to the 'caseness' scale from the Stirling County study (13) and 'impairment' scale from the Midtown study (17) as well as diagnostic impressions. They include also theoretically and actuarially derived objective measures of different types of disorder.

Preliminary results

As was mentioned earlier, our analyses of the data are far from complete at this writing. They have been sufficient, however, for us to select some results that will illustrate the range and variety of our data and a few of the intriguing problems that they raise.

SOME OVER-ALL CONTRASTS AMONG PATIENTS, LEADERS, AND COMMUNITY SAMPLE RESPONDENTS

The way they see themselves

We can get some idea of the differences in the way respondents in our patient and non-patient groups see themselves from their responses to the Self-administered Questionnaire that was filled out toward the end of both the PSS and the SIS interview by all respondents. On the Self-administered Questionnaire, positive descriptions of behaviour were placed with negative descriptions and the respondents were asked to rank-order the behaviours from the one that was most like them to the one that was least like them.

By and large, respondents from the non-patient groups are much more likely to report seeing themselves in positive terms than respondents from the patient groups. For example, on the last page of the Self-administered Questionnaire, we presented three sets of alternative descriptions. In each set, two descriptions were of negative behaviour and one was positive. The three sets of descriptions are as follows:

1. (*a*) A person who is relaxed, arranges work to get important things done first, and does not spend too much time on details; does not have any fixed plan, but deals with each situation as it arises. Once the problem has been dealt with, this person likes to forget it and go on to the next.

(*b*) A person who worries a lot about little things; seems to be moody and unhappy all the time; and cannot sleep at night brooding about the past, and worrying about things that *might* go wrong.

(*c*) A person who just cannot leave the house without going back to see whether the gas stove was left lit or not; always goes back again just to make sure the door is locked, and is afraid to ride up and down in elevators.

2. (*a*) A person who is very quiet; does not talk much to anyone; acts afraid of people; stays alone and daydreams all the time; and shows no interest in anything or anybody.

(*b*) A person who is very suspicious; does not trust anybody; and is sure that everybody is against this person.

(*c*) A person who is open and says what is on his or her mind; gets along easily with people, including strangers; is always willing to listen to people's problems and help them out when they are in difficulty; and sometimes makes promises that are hard to keep, but tries to keep them. Once in a while this person feels let down, as when somebody does not repay a loan, but the let-down feeling does not last for long.

3. (*a*) A person who drinks too much; goes on a spree when there is money in the pocket. This person promises to stop drinking, but always goes off again.

(*b*) A person who has been telling lies for a long time now, and stealing things. Others are very upset about these acts, but the person pays no attention to others.

(*c*) A person who thinks of himself or herself as serious and respectable; works hard, and would be shocked if someone suggested doing anything dishonest; saves money and does not buy things on credit; does not believe in wasting time in idle conversation or parties. Some people find this person a little too serious

As Table 4.6 shows, the patients are far more likely than the non-patients to see a resemblance to themselves in the negative descriptions. More than half the clinic patients choose at least one negative self-description, as do three-quarters of the

TABLE 4.6. *Number of negative self-descriptions on three sets of items at end of self-administered questionnaire by leader, community sample, clinic patient, prisoner, and hospital patient status (percentage)*

Number of negative self-descriptions	*Respondent status* *Leader*	*Community sample*	*Clinic patient*	*Prisoner*	*Hospital patient*
0	92·2	68·0	43·5	28·6	25·0
1	6·3	24·9	29·6	19·0	43·8
2	1·6	6·6	21·3	47·6	18·8
3	—	0·4	5·6	4·8	12·5
Total %	100·1	99·9	100·0	100·0	100·1
Base for %	(64)	(241)	(108)	(21)	(32)

Note. With scores dichotomized at 0 versus 1 or more, and with clinic patients, prisoners, and hospital patients combined, chi-square tests show the following: that the probability is less than 0·01 that the over-all results could have occurred by chance; that the leaders are significantly less likely (at the 0·01 level) than the community sample respondents to give negative self-descriptions; and that the community sample respondents are significantly less likely (at the 0·01 level) than the combined patients and prisoners to give negative self-descriptions.

prisoners and hospital patients; by contrast, a large majority of the community respondents, and an even larger majority of the leaders present self-descriptions that are uniformly positive on these items.

The kinds of things that happen to them

After the psychiatrist-interviewers made their judgemental evaluations and after the respondents had filled out the Self-administered Questionnaire, the interview continued with a final section that was also the same for both the SIS and the PSS. This final section included, along with background data, the Stressor Inventory designed to elicit stressful objective events experienced by the respondents.

The first question on the Stressor Inventory went like this: 'What was the last major event in your life that, for better or for worse, interrupted or changed your usual activities?' (Probe: 'For example, events affecting your occupation, your physical health, your living arrangements, your relations with other family members, your friends, or your personal values or beliefs.') In answer to this question, respondents (excluding in-patients and prisoners because of the influence of the special circumstances of their institutionalization) reported seventy-seven different types of events. Two coders, working independently, were able reliably to classify these types of events into those that they thought most people would view as mainly positive, e.g., getting married; mainly negative, e.g., death of a loved one; or indeterminate, e.g., a change of job without evidence as to the quality of the change. Of the seventy-seven types of events reported, there was disagreement between the coders on only four which were accordingly assigned to the 'indeterminate' category.

The majority of types of events reported, 64 per cent, were negative in the above sense; 26 per cent were positive; and the remaining 10 per cent were indeterminate.

As Table 4.7 shows, all but small minorities of the respondents from patient and non-patient groups alike could recall events that, for them, fit our definition of a stressor. The most striking result, however, is that only for the leaders were the events recalled more often positive than negative. For the majority of the patients and for the majority of the respondents in the

TABLE 4.7. *Positive or negative quality of last major stressor recalled by leaders, community sample, and out-patients (percentage)*

Stressor	*Leader*	*Respondent status* *Community sample*	*Out-patient*
None recalled	7·6	16·5	21·9
Positive event	40·9	22·0	18·4
Negative event	34·8	53·5	55·3
Quality of event indeterminate	16·7	7·9	4·4
Total %	100·0	99·9	100·0
Total respondents	(66)	(254)	(114)

Note. Chi-square tests show that: the probability is less than 0·01 that these over-all results could have occurred by chance; that leaders report significantly (at the 0·05 level or better) more positive events (or significantly fewer negative events) than either patients or community sample respondents; that there is no difference that approaches statistical significance between the community sample and the patients.

community samples, by contrast, the stressors reported were negative.

Further analysis will tell us to what extent this difference among leaders and all other respondents is mainly a class phenomenon of, for example, getting better jobs, moving to better neighbourhoods—a difference between stressors that are related to greater opportunity as opposed to stressors related to less opportunity such as loss of job or forced relocation.

The symptoms they report

Symptom items on the SIS and PSS schedules are divided into two main types: first, those designed to elicit symptoms of mental status—that is, disturbances of cognition, effect, and volition; second, those designed to elicit evidence of disability in role-functioning. Even our preliminary analyses with fairly primitive indexes constructed from these symptom data show marked contrasts between the leader, community sample, and patient groups. Typically, as would be expected, the leaders score lowest on these indexes and the patients highest. Let us consider the SIS items on mental status by way of illustration.

There are 46 such symptom items in the SIS interview schedule, including the Midtown 22-item screening instrument as a subset (12). With their responses scored as in the Midtown screening instrument, our community sample respondents averaged 3·17 of these 22 symptoms by contrast with 2·83 in the Midtown study sample (12). About 8 of these 22 Midtown study screening items overlap with some of the 24 items used in the Stirling County study, and we added 3 more that were used in the Stirling County study questionnaire. Summary descriptions of the total 46 mental status items in the SIS, with designations of whether they were used in the Midtown and/or Stirling County studies are contained in Table 4.8. In the following comparisons of the patient and non-patient groups in our study, we shall omit at first the prisoners. They will be considered separately later on.

As would be expected, the median number of symptoms shown in Table 4.9 is much higher in the patient than in the non-patient groups. For the leaders, in fact, there is no overlap at all between their interquartile range and that of either of the patient groups on the over-all 46 item scores, the portion of them made up of the 22-item screening instrument from the Midtown study, or the 24 items that we added. Note also in Table 4.9 that on the 22-item Midtown measure the out-patients tend to score higher than the in-patients. We would hardly expect this to happen if the Midtown measure provided an adequate index of the relative severity of disorder in the different groups.

Ignoring this difficulty for the moment, let us ask whether large proportions of the leaders and community sample respondents actually do resemble the patient groups in quantity of symptoms—as the high rates of 'cases' reported in the Midtown study (23·4 per cent) (17, pp. 136–8, 333) and Stirling County study (slightly over 50 per cent) (13, p. 356) would suggest. In doing so, we shall take as evidence of such similarity an individual score at or above the lowest score marking the interquartile range for the clinic patients as shown in Table 4.9. We chose out-patients as the standard here because, in the Midtown study which reported a lower rate than the Stirling County study, the large majority of the 'cases' in the community were held to resemble out-patients rather than in-patients.

Table 4.10 shows that the proportions of the community

TABLE 4.8. *Brief descriptions of 46 mental status items from SIS ('M' indicates that symptom was included in Midtown 22-item screening instrument; 'S' indicates item was included in Stirling County study questionnaire)*

M & S	1.	Heart beats hard (often)
M	2.	Shortness of breath (often)
M	3.	Nervous (often)
M & S	4.	Trouble getting to sleep (often)
M & S	5.	Cold sweats (often)
M & S	6.	Hands tremble (often)
M & S	7.	Headaches (often)
M	8.	Fainting (more than a few times)
M & S	9.	Appetite (poor)
M & S	10.	Spirits (low or very low)
M	11.	Feel weak all over
M	12.	Cannot get going
M	13.	Hot all over
M	14.	Restlessness
M & S	15.	Sour stomach
M	16.	Memory not all right
M	17.	Clogging in nose
M	18.	Personal worries get one down physically
M	19.	Nothing turns out right
M	20.	Wonder if anything is worthwhile
M	21.	Feel somewhat apart
M	22.	Worrying type
S	23.	Nightmares
S	24.	All kinds of ailments
S	25.	Physical health (poor, very poor)
	26.	Take sleeping pills or other drugs
	27.	Misunderstood
	28.	'High' periods cause concentration problem
	29.	Drinking: trouble with health and work
	30.	Make people afraid for fun of it
	31.	Anxiety
	32.	Trust nobody
	33.	Mind not working
	34.	Voices
	35.	Drinking (arguments)
	36.	Worried about sex
	37.	See things
	*38.	People talk behind back
	39.	Quick tempered
	40.	People object to companions
	41.	Special thoughts
	42.	Wish dead
	43.	Feel blue
	44.	Special fears
	45.	Special habits
	46.	Other problems

* Included in Midtown study but not among those in Midtown 22-item screening instrument.

TABLE 4.9. *Median number of symptoms and interquartile range on 22 Midtown items, 24 additional items, and total 46 items, by leaders, community sample, clinic and hospital patients: SIS respondents only*

	Type of respondent			
Symptom sets	*Leaders*	*Community sample*	*Clinic patients*	*Hospital patients*
22 Midtown items				
Median	0·82	1·94	6·42	5·0
Interquartile range	(2·9–0·10)	(5·28–0·44)	(10·88–4·05)	(8·17–3·0)
24 Additional items				
Median	1·00	2·07	6·60	6·83
Interquartile range	(2·88–0·14)	(4·42–0·65)	(10·06–4·81)	(8·75–4·0)
46 Item total				
Median	1·8	4·5	13·33	11·0
Interquartile range	(5·42–0·43)	(9·50–2·28)	(19·63–8·08)	(14·5–7·0)
Total respondents	(41)	(124)	(59)	(24)

sample who by this measure quantitatively resemble the out-patients are substantial, especially on the 22-item Midtown measure, where over one-third of the community sample score above the lower quartile of clinic scores. Note, however, a rather striking set of trends bearing on our question. The percentages of both the leader and especially the community sample respondents quantitatively resembling the majority of out-patients are larger on the 22-item Midtown measure than on the 24 additional items. The trends for the hospital patients, by contrast, are just the opposite, with the proportions quantitatively resembling the out-patients tending to be larger on the 24-item set. Together with the results shown in Table 4.9 what this suggests, of course, is that the proportion of severe symptomatology is greater among the 24 items that we added than among the 22-item screening instrument from the Midtown study. Not in the 46 items as a whole nor in the two subscales, however, is there an adequate reflection of the greater severity of disorder among the in-patients than among the out-patients.

TABLE 4.10. *Percentage at or above 25th centile of clinic scores on 22 Midtown items, 24 additional items, and total 46 item symptom measures, by leaders, community sample, clinic and hospital patients*

Scores at or above 25th centile for clinic patients on each symptom set		*Type of respondent*		
	Leaders	*Community sample*	*Clinic patients*	*Hospital patients*
Scores of 4 or more on 22 Midtown items	14·6	34·7	81·4	66·7
Scores of 5 or more on additional 24 items	12·2	24·2	79·7	70·8
Scores of 8 or more on total 46 items	17·1	34·7	78·0	79·2
Base for %	(41)	(124)	(59)	(24)

Note: Chi-square tests show that: with the clinic and hospital patients combined, the probabilities are less than 0·01 that the over-all results with each of the three symptom sets could have occurred by chance; the differences between the leaders and the community sample and between the community sample and each of the patient groups are significant at the 0·05 level or better.

The most likely reason is that the ratio of more severe to less severe symptoms in all three measures is too low to permit the measures to reflect the greater severity of disorder in the in-patient group. If this is so, sheer quantitative resemblance to patients on a screening instrument reflecting such imbalance of more severe *v.* less severe symptoms is an inadequate measure of disorder in the community population; the proportion of 'cases' in the general population thus measured will vary directly with the proportion of less severe symptoms in the screening instrument.

We mentioned earlier that the SIS prisoner group would be considered separately on these measures. Table 4.11 shows why. The prisoners seem almost a group apart, scoring higher than out-patients or in-patients on all three measures. We think that this may be because the situational stress of their recent imprisonment has interacted with whatever pathology previously existed to produce high scores on both the more and the less serious symptom items among the 46 we used. More analysis is obviously needed, however, of not only these but a larger

TABLE 4.11. *Prisoners' median number of symptoms, interquartile range, and percentage above lower quartile of clinic scores on 22 Midtown items, 24 additional items, and total 46 item symptom measure: SIS respondents only (n = 11 prisoners)*

	Symptom measure		
Summary statistic	*22 Midtown items*	*24 additional items*	*46 items total*
Median	9	11	18
Interquartile range	4·5–12·5	4–13	10·5–24·5
Percentage above lower quartile of clinic scores	81·8	81·8	81·8

sample of prisoners and of the effects of situational stress on these symptom scores.

The way they are seen by the psychiatrists who interviewed them

At that point toward the end of the interview with either the PSS or the SIS schedule where respondents filled out the Self-administered Questionnaire, the psychiatrists who were interviewing them took the opportunity to record their clinical judgements: the Stirling County study rating on a scale of 'caseness', the Midtown study rating on a scale of 'impairment', and a diagnostic impression. Our analyses of the diagnostic impressions are not far enough along to report. We would like to note in passing, however, that 4·5 per cent of the PSS respondents and 1·6 per cent of the SIS respondents in the community sample were diagnosed as schizophrenic (3·1 per cent over the whole community sample); only half of these subjects reported in response to direct questioning that they had ever been in treatment for their disorder with members of the mental health professions. We mention this because it contradicts the often-heard speculation that almost all schizophrenics —at least in modern societies—are likely to have a treatment history.

Our focus in this report, however, will be on the Stirling 'caseness' ratings and the Midtown 'impairment' ratings. Consider first the results shown in Table 4.12 on the 'caseness' ratings made by the psychiatrists on the basis of actually seeing the respondents (Interviewer Caseness Rating) and regardless

TABLE 4.12. *Interviewer 'caseness' ratings according to leader, community ethnic sample, or patient status (percentage; structured interview schedule respondents and psychiatric status schedule respondents combined)*

	Status of respondent				
'Caseness' rating	*Leaders*	*Community sample*	*Out-patients*	*Prisoners*	*In-patients*
A = Almost certainly psychiatric case	11·9	19·8	76·3	79·2	97·9
B = Probably a psychiatric case (too vague to rate 'A')	7·5	10·5	7·6	8·3	2·1
C = Possibly psychiatric case (border-line)	25·4	35·4	11·9	8·3	—
D = No evidence of symptoms. Some evidence of good health and adjustment	55·2	34·2	4·2	4·2	—
Total %	100·0	99·9	100·0	100·0	100·0
Total respondents	(67)	(257)	(118)	(24)	(48)

Note. With the out-patients, prisoners, and hospital patients combined, chi-square tests show that: the probability is less than 0·01 that these over-all results could have occurred by chance; the leaders and community sample respondents both have significantly fewer As and significantly more Cs and significantly more Ds than the combined patients and prisoners; the leaders have significantly more Ds than the community sample respondents. Each of the specific differences mentioned above is significant at the 0·05 level.

of which interview schedule (SIS or PSS) was used. Clearly, and as would be expected, there is sharp contrast; the patients are rated far sicker than the community respondents who, in turn, register more disorder than the leaders.

Consider now the Midtown rating of 'impairment,' that study's main measure of a 'case.' Table 4.13 shows that there is also marked contrast among the leader, community sample, and patient groups.

Note that the estimate of impaired in the community ethnic samples, 27·6 per cent, is slightly higher than the 23·4 per cent reported for Midtown. By contrast, the Stirling County researchers saw considerably more illness according to their 'caseness' rating in the rural maritime county they studied than our psychiatrists saw in the community ethnic samples from Washington Heights. The Stirling County figures were: A = 30·5 per cent; B = 24·6 per cent; C = 26·2 per cent; and D = 18·7 per cent (13, p. 121). To these ABCD ratings, the Leightons attach the following average 'subjective probabilities' that the person rated would actually be a case if given a 'full diagnostic investigation': A = 0·9; B = 0·7; C = 0·4; and D = 0·1. Thus, if we dichotomize the ratings into A and B as more likely

TABLE 4.13. *Interviewer 'impairment' ratings according to leader, community ethnic sample, or patient status (percentage; respondents to structured interview schedule and psychiatric status schedule combined)*

	Status of respondent				
*Impairment ratings**	*Leaders*	*Community sample*	*Out-patients*	*Prisoners*	*In-patients*
Unimpaired					
0. No significant symptom formation (symptom free) = 'Well'	43·3	25·3	1·7	4·5	—
1. Mild symptom formation, but functioning adequately = 'Mild'	29·9	29·2	6·0	4·5	—
2. Moderate symptom formation with no apparent interference in life adjustment = 'Moderate'	16·4	17·9	9·4	—	—
Total % unimpaired	*89·6*	*72·4*	*17·1*	*9·0*	—
Impaired					
3. Moderate symptom formation with some interference in life adjustment = 'Marked'	7·5	18·3	20·5	9·1	2·1
4. Serious symptom formation functioning with some difficulty = 'Severe'	3·0	6·6	36·8	18·2	12·5

TABLE 4.13—*cont.*

5. Serious symptom formation and functioning with great difficulty = 'Nearly incapacitated'	—	2·7	22·2	45·5	47·9
6. Seriously incapacitated, unable to function = 'Incapacitated'	—	—	3·4	18·2	37·5
Total % impaired	*10·5*	*27·6*	*82·9*	*91·0*	*100·0*
Total percentage	100·1	100·0	100·0	100·0	100·0
Total respondents	(67)	(257)	(117)†	(22)‡	(48)

* The terms in 'quotes' are the summary descriptions used by the Midtown researchers in reporting ratings obtained in terms of the more detailed descriptions listed.

† Excludes one respondent for whom no rating was made.

‡ Excludes two respondents for whom no rating was made.

Note. With ratings dichotomized into impaired (3, 4, 5, or 6) and unimpaired (0, 1, 2) and with out-patients, prisoners, and in-patients combined, chi-square tests show the following: that the probability is less than 0·01 that the over-all results could have occurred by chance; that the leaders are significantly (at the 0·05 level) less impaired than the community sample respondents; and that the community sample respondents are significantly (at the 0·01 level) less impaired than the combined patients and prisoners.

to be cases versus C and D as less likely to be cases, 'caseness' ratings for the Washington Heights community sample would indicate a rate of about 30 per cent as opposed to about 55 per cent for Stirling County (13).

There is an interesting question here as to why the Washington Heights community ethnic samples should appear slightly sicker than the Midtown sample by the 'impairment' rating while at the same time appearing much healthier than the Stirling County sample by the 'caseness' rating. There is no apparent substantive explanation for this in terms of differences among the three populations studied. Later on, as the reader will see, the difference, especially on the Stirling 'caseness' rating, may be due in part to the fact that the psychiatrists on the Stirling County study made their ratings not on the basis of a face-to-face interview, but rather on the basis of a written record of the interview.[1]

Meanwhile, note that the respondents rated in any one of the 'impaired' categories were thought by the Midtown study psychiatrists to suffer disorder analogous to the clinical conditions seen in psychiatric patients. Those rated 'marked' and 'severe' were held to be comparable to patients in ambulatory treatment; those rated in either of the two 'incapacitated' categories, as comparable to the clinically hospitalized (17, pp. 136–8, 333). It is therefore startling to find, as Table 4.13 shows, that the group of 'impaired' among the respondents from the community ethnic samples *are not*, on the average, comparable to the groups of impaired in any of the patient groups. Typically, the impaired community member was judged to show 'moderate symptom formation with some interference in life adjustment' by our psychiatrists. (The Midtown psychiatrists also placed the majority of the respondents they rated impaired in this category.) By contrast, the large majority of the out-patients who were rated impaired were placed in one of the categories involving 'Serious symptom formation . . .'—that is, 'severe' or higher. Some arithmetic with the results in Table 4.13 shows that only 33·8 per cent of the community sample impaired were placed in the 'serious symptom formation' categories as against 75·5 per cent of the out-patient impaired—a difference that chi-square test shows could have occurred by chance less than one time in a hundred. According to the Midtown impairment rating as applied by our interviewing psychiatrists, then, a *typical* 'case' in the general population is simply not the same as a *typical* 'case' in the clinic much less the mental hospital.

Moreover, when the impairment ratings of respondents in the community ethnic samples and the out-patient sample who were rated 'A', the most extreme category of illness on the Stirling scale, are examined, we find that only 41 per cent of the community respondents who were rated 'A', received Midtown impairment ratings as high as 'severe'. By contrast, fully 77

---

[1] The Stirling County study investigators attempted to deal with past as well as current 'caseness' whereas we attempted to deal only with current 'caseness'. However, the Stirling County researchers estimated that 90 per cent of their 'cases' were current. Thus their inclusion of past cases can hardly account for the difference.

per cent of the out-patients rated 'A' were also rated 'severe' or higher. Chi-square test indicates that the probability is less than 0·01 that this difference could have occurred by chance. It would seem, therefore, that by the Stirling caseness rating as well as by the Midtown impairment rating, the typical 'case' in the community is quite different from the typical 'case' in the psychiatric clinic or hospital—at least, as we said earlier, as these ratings have been interpreted by the interviewing psychiatrists.

SOME CONTRASTS IN THE RESULTS OBTAINED WITH THE PSS *v.* THE SIS

Thus far, we have presented results that show sharp differences between the leader, community sample, and patient respondents. Even at this early stage of the analysis, however, we know that there are differences in the psychiatric pictures drawn by the two main data collection instruments—the PSS and the SIS. Moreover, these differences are far more pronounced in the non-patient than in the patient groups. Our main clues to these instrument contrasts thus far are in the ratings the psychiatrists make of respondents depending on whether the SIS or the PSS was used in the interview.

PSS v. *SIS contrasts among community sample respondents from different ethnic groups*

When we look at the interviewer 'caseness' ratings on the Stirling County scale made of the different ethnic groups in the community samples, we see in Table 4.14 that they appear to contrast sharply depending upon which schedule the psychiatrist used. The Puerto Ricans are much more likely to be seen as cases when interviewed with the SIS than with the PSS. To a lesser extent, so are all the other ethnic groups except the Irish, where there is a tendency for the reverse to be true. As Table 4.15 shows, with the Midtown impairment rating as the measure of disorder, the differences remain sharp for Puerto Ricans while decreasing for the other groups.

The chi-square tests indicate more than minimal possibilities that these differences according to type of interview could have occurred by chance. Nevertheless, the magnitude of the SIS versus PSS differences in some instances and the consistency

TABLE 4.14. *'Caseness' in community ethnic samples as judged by interviewer according to type of interview schedule used (percentage; 'case' = A or B on Stirling caseness rating)*

	Ethnicity					
Type of interview schedule	*White Protestant of old American ancestry*	*Jewish*	*Irish*	*Negro*	*Puerto Rican*	*Total*
SIS	31·3	34·6	23·1	31·4	57·1	34·7
Base for %	(16)	(26)	(26)	(35)	(21)	(124)
PSS	12·5	15·4	38·7	27·6	29·0	26·3
Base for %	(16)	(26)	(31)	(29)	(31)	(133)
Statistical tests:						
χ^2	0·73	1·64	0·60	—	3·03	2·12
d.f.	1	1	1	1	1	1
p	> 0·30	< 0·20	> 0·30	> 0·30	< 0·10	< 0·20

TABLE 4.15. *'Impairment' in community ethnic samples as judged by interviewer according to type of interview schedule used (percentage; 'case' = 3, 4, 5, or 6 on Midtown impairment rating)*

	Ethnicity					
Type of interview schedule	*White Protestants of old American ancestry*	*Jewish*	*Irish*	*Negro*	*Puerto Rican*	*Total*
SIS	31·3	23·1	26·9	25·7	52·4	30·6
Base for %	(16)	(26)	(26)	(35)	(21)	(124)
PSS	25·0	15·4	25·8	31·0	25·8	24·8
Base for %	(16)	(26)	(31)	(29)	(31)	(133)
Statistical tests:						
χ^2		0·13			2·75	0·82
d.f.		1			1	1
p		> 0·30			< 0·10	>0·30

with which the higher rate is found for SIS respondents (8 out of 10 comparisons in Tables 4.14 and 4.15) lead us to regard the schedule differences for the community respondents as important points of departure for further investigation. This belief is further reinforced by the fact that the schedule differences shown above are increased when community sample members who reported being or having been in treatment with members of the mental health professions are removed from Tables 4.14 and 4.15. Such respondents were more likely than others to be cases according to the Midtown and Stirling ratings, and 15 out of 25 of them happened to be interviewed with the PSS schedule.

The class relationship

As was mentioned earlier, the most consistent finding reported in studies of psychiatric epidemiology is an inverse relation between disorder and social class. In the present study, the findings so far as the judgemental ratings on the Midtown and Stirling County study scales are concerned are consistent with previous research. Thus, Table 4.16 shows that the trend among

TABLE 4.16. *'Cases' among community sample respondents as judged by interviewer according to family income of respondents: SIS versus PSS respondents separately (percentage)*

Definition of 'case'	*Less than $5,000*	*Family income* *$5,000–9,999*	*$10,000 or more*
		SIS respondents	
A or B on Stirling caseness rating	51·1	27·7	22·6
Base for %	(45)	(47)	(31)
3, 4, 5, or 6 on Midtown impairment rating	40·0	25·5	25·8
Base for %	(45)	(47)	(31)
		PSS respondents	
A or B on Stirling caseness rating	35·0	24·2	19·4
Base for %	(40)	(62)	(31)
3, 4, 5, or 6 on Midtown impairment rating	37·5	21·0	16·1
Base for %	(40)	(62)	(31)

Note. Chi-square tests show that the probability is: less than 0·025 that the SIS results in the Stirling rating could have occurred by chance; and less than 0·10 that the PSS results on the Midtown rating could have occurred by chance. Inspection shows, moreover, that the trends are consistent in all four sets of results.

respondents from the community ethnic samples is for the higher rates of disorder as measured by the interviewer caseness and impairment ratings to occur at the lower levels of family income. By and large this holds regardless of which interview schedule, the SIS or PSS, was used. The results are similar when the educational level of the household head is substituted for family income as the indicator of class. For this report, we are omitting analysis of relations between judged disorder and the third major indicator of class, occupational level, because we have not as yet conducted the necessary preliminary analyses of our data on occupations per se that will help us to order the occupations into levels.

We have been describing thus far results obtained with judgemental 'caseness' and 'impairment' ratings made by the psychiatrist on the basis of actually seeing the respondent. These ratings are not blind to the class status of the respondents. Was there, then, a bias toward rating lower-class respondents sicker than higher-class respondents (10)? In part to explore this possibility, we asked four of the psychiatrists who did interviewing on the study to also rate written records of a subsample of 168 interviews, for the most part excluding their own, after information on the status of the respondents had been removed. Of this subsample, 84 were respondents from the community ethnic samples.

Table 4.17 shows the results for the subsample of SIS respondents from the community ethnic samples. Other than a tendency for the reviewer to see somewhat more respondents as cases on the Stirling rating than the interviewer, there is no marked difference. It seems unlikely that the blind reviews would alter the class relationship based on the ratings by the interviewer for these respondents, although there is a suggestion of this possibility in the fact that of the 16 SIS leaders (not shown in Table 4.17) who were in the sample receiving blind ratings, 43·8 per cent were rated A or B by the reviewer in contrast to only 18·8 per cent by the original interviewer.

For the PSS respondents, the picture is different. As Table 4.18 shows, the reviewers apparently see the higher-income PSS members of this subsample of community respondents much differently from the way they were seen by the interviewer. The reviewers are much more likely to see such respondents, especi-

TABLE 4.17. *'Cases' among subsample of community respondents as judged by interviewer versus 'blind' reviewer (trained in the same manner) according to family income: SIS respondents only (percentage)*

Definition of a 'case' by type of rater	*Family income* Less than $5,000	$5,000–9,999	$10,000 or more	*Total**
A or B on Stirling caseness rating:				
Interviewer	60·0	26·7	18·2	36·6
'Blind' reviewer	60·0	40·0	27·3	46·3
Total respondents rated by both	(15)	(15)	(11)	(41)
3, 4, 5, or 6 on Midtown impairment rating:				
Interviewer	33·3	20·0	27·3	26·8
'Blind' reviewer	46·7	20·0	9·1	26·8
Total respondents rated by both	(15)	(15)	(11)	(41)

* Further analysis shows that the probability is 0·14 that the reviewer over interviewer gain in total cases according to the Stirling rating could have occurred by chance. The test used is that cited in W. A. Wallis and H. V. Roberts, *Statistics: A New Approach* (New York: The Free Press, 1956), p. 433, for differences between two measures each with scores of 0–1 on the same sample.

ally those with incomes of $10,000 or more, as cases. This was true also for the 9 PSS leaders who received blind reviews; for example, none of the 9 was seen as impaired by the interviewer whereas 4 of these leaders were judged impaired by the reviewers.

These startling results were obtained with the first group of 'blind' reviewers who had been trained by our study psychiatrists in the same way as the original interviewers—that is, on the basis of published accounts of the Midtown study and Stirling County study rating procedures. These interviewer versus reviewer differences turned out to be substantially the same when the ratings on a somewhat smaller sample of respondents were made by psychiatrists who were trained by Midtown study and Stirling County staff to be reliable with the psychiatrists who did the ratings on these two studies.

TABLE 4.18. *'Cases' among subsample of community respondents as judged by interviewer versus 'blind' reviewer (trained in the same manner) according to family income: PSS respondents only (percentage)*

Definition of a 'case' by type of rater	*Family income* *Less than $5,000*	*$5,000–9,999*	*$10,000 or more*	*Total**
A or B on Stirling caseness rating:				
Interviewer	18·2	15·0	8·3	14·0
'Blind' reviewer	18·2	45·0	75·0	46·5
Total respondents rated by both	(11)	(20)	(12)	(43)
3, 4, 5, or 6 on Midtown impairment rating:				
Interviewer	18·2	5·0	0·0	7·0
'Blind' reviewer	9·1	25·0	50·0	27·9
Total respondents rated by both	(11)	(20)	(12)	(43)

* Further analysis shows that the reviewer over interviewer gains on both the Stirling and Midtown ratings in total cases are significant at the 0·01 level. The test used is that cited in W. A. Wallis and H. V. Roberts, *Statistics: A New Approach* (New York: The Free Press, 1956), p. 433, for differences between two measures each with scores of 0 or 1 on the same sample.

What could possibly account for this contrast, and what are its implications? Is it possible, for example, that the PSS is more likely than the SIS to elicit symptomatology in high status groups? If so, why do the interviewers who actually see the respondents tend to discount these symptoms in higher status respondents while the reviewers, blind to this status, do not? Is it that clues to positive functioning present in the interview situation but absent from the paper record of the PSS influence the interviewer? Or does outright bias in favour of the manner and style of the high status respondent lead the interviewer to ignore the symptomatology reported?

And what of the SIS which shows little or no such difference on review? Does the SIS miss the symptomatology of high status respondents? If so, is this not also bias of a different sort—

resulting perhaps from some of the types of symptoms included in the SIS (a number of them psychophysiological) (1). Or is there a problem in its use of closed, fixed alternative questions? Do such features of the SIS magnify distress in low status groups while missing the expression of distress by respondents of higher status (11)? These are major questions for further analysis, including analysis of the tape-recordings of these interviews.

PSYCHIATRISTS *v.* LAY INTERVIEWERS

The procedure of having psychiatrists conduct the interviews in the present study was expensive. The field work under these circumstances, for example, took two years and four months to complete. An important question arises, therefore, as to whether the psychiatrists made a difference, as compared with lay interviewers, in the data collected even on the SIS which required far less than the PSS in the way of clinical background.

One way in which we have approached this question is by analysis of a panel of community sample respondents of whom symptom data have been obtained at three points in time over a five-year period. This panel consists of 41 respondents who had been interviewed in two of our previous surveys as well as in the present study. In each of these interviews, there was a common core of questions from the 22-item screening instrument developed in the Midtown study. Lay interviewers at time 2 were matched in ethnicity with Negro and Puerto Rican respondents but not at time 1; at time 3 the interviewing was done by psychiatrists, also unmatched with their respondents in ethnic background. Individual respondents' scores in each survey were computed and their three scores compared. The mean number of symptoms at time 1 is 2·02; at time 2, 1·83; and at time 3, 2·88. Of the 22 symptoms, 13 are reported more frequently to the psychiatrists than to either of the two previous interviewers; 18 are reported more frequently to the psychiatrists than in the previous survey (time 2), and a sign test shows that with 1 tie this difference is significant at the 0·1 level; 15 symptoms are reported more frequently to the psychiatrists than on the first interview (time 1), and again a sign test shows that with 5 ties this difference is significant at the 0·1 level.

It would appear, then, that there is an increase in the facility with which respondents admit symptoms—at least of the types asked about in the 22 items—when interviewed by a psychiatrist, as compared to what they will admit to a lay interviewer, even if matched by ethnicity with the respondent. The psychiatrists were identified as physicians to respondents before the interview. An interesting question arises as to whether and to what extent it was their formal status that elicited greater admission of symptoms or their greater skill in handling the interview.

Summary and implications

This report has focused on the central unsolved methodological problem in psychiatric epidemiology, the problem of how to conceptualize and measure psychiatric disorder independently of treatment status in general populations. The project we have described is bringing to bear on this problem data obtained from two different types of research interviews (the PSS and the SIS) conducted in the community by 15 psychiatrists.

The subjects of the research, over 500 to date, have come mainly from five different ethnic groups, including Negroes and Puerto Ricans, in the Washington Heights section of New York City. About 190 are psychiatric patients; 257 are members of community samples; and 67 are community leaders. These subjects have been randomly alternated between the two contrasting interview instruments, thereby permitting controlled comparisons of the results obtained by each. The large majority, 84 per cent, of the interviews have been tape-recorded as well as scored on the interview forms by the psychiatrist interviewers. Both judgemental measures of disorder and objective measures of disorder are being developed to compare the three main groups of subjects and to investigate differences in results obtained by the two interview instruments. Since a number of the subjects of the present study had been interviewed during the preceding 5 years by lay interviewers, i.e., non-clinician, non-MD, and since a subset of symptom items has been carried over from these previous studies, it has also been possible to compare some of the results obtained by psychiatrists with results obtained by lay interviewers.

Although our analyses of the data are incomplete, some of the results obtained so far have important implications. Especially worthy of note are the following findings:

(1) Both judgemental measures of disorder and preliminary objective measures of disorder discriminate sharply between leaders, patients, and community sample respondents.

(2) As judged according to Stirling County study and Midtown study rating procedures, typical 'cases' in the community are not the same as typical 'cases' in the psychiatric clinic or hospital.

(3) Data from the two different types of research interview (PSS and SIS) present different pictures to psychiatrists of the psychiatric conditions of the community respondents. For example, relations between judged disorder and ethnic status and class status vary with type of research interview and with whether the judgements were made on the basis of the interview or were 'blind' reviews of the paper record of the interviews.

(4) Community respondents appear more seriously ill to psychiatrists on the basis of a written record from which direct clues to social and patient status have been removed than they appear in a face-to-face interview.

(5) Community respondents admit more symptoms to psychiatrists than they admitted to lay interviewers in previous interviews.

The implications of these findings are important first as they bear on key sets of problems involved in the conceptualization and measurement of psychiatric disorder in general populations; and second, as they influence our next steps toward assessing the relative importance of social causation *v.* social selection factors in social class differences in rates of psychopathology.

Implications for key problems in the conceptualization and measurement of psychiatric disorder

One of the key sets of problems in measuring psychiatric disorder in general populations centres on subcultural differences in modes of expressing distress (4, chapter 6). The class and ethnic differences we found between the two instruments suggest that such subcultural differences may be related not only to the types of symptoms asked about, but also to the types of questioning procedures that we used to ask about symptoms.

Another of the key sets of problems is concerned with relations between mental status symptoms and evidence of disability in role-functioning.[1] Here, the differences in results obtained in face-to-face interviews in comparison with 'blind' reviews may add a major dimension for further investigation, since they suggest that positive strengths are being missed in the scoring systems of the interviews.

With regard to the question of positive functioning, it is interesting to recall our striking finding, in using Midtown study and Stirling County study rating procedures, that a typical 'case' in the clinic or hospital is not the same as a typical 'case' in the community. We can now see as complementing this finding the fact that our interviewer *v.* reviewer differences were far less pronounced for patients than for community respondents. We would expect just such a result if the patients showed less evidence of positive functioning than the community respondents in the interview situation. The plausibility of this explanation is enhanced by the fact that we find sharp interviewer *v.* reviewer differences for the leaders; we cannot, therefore, dismiss the relative lack of such a difference for the patients as just a matter of greater convergence of all ratings at the extremes of ability and disability.

Implications for long-term goal of investigating the relative importance of social causation v. *social selection factors in class differences in rates of psychopathology*

Let us turn now to the implications of these findings for our long term goal of investigating the major substantive issue posed by the frequently reported finding of an inverse relationship between social class and psychiatric disorder—over-all rates as well as the subtypes, schizophrenia and personality disorder (4, chapter 2). Accepted at face value, the central issue raised by

---

[1] See E. Ginzberg, J. K. Anderson, S. W. Ginsberg, and J. L. Herma, *The Lost Divisions* (New York: Colombia University Press, 1959), especially pp. 167–93. Note, moreover, that although both the Midtown study and, to a lesser extent, the Stirling County study researchers gave central importance to measures of impairment of functioning, the data on which these measures are based are weak. As the investigators in both studies acknowledge, the full importance in getting good data on role-functioning became evident to them only after their data had been collected, see (17), p. 57, and (13), p. 53.

these findings is the one so vividly posed by Faris and Dunham and their critics thirty years ago (8). Is low social status more a cause or is it more a consequence of psychiatric disorder? On the basis of research to date, it has been impossible to tell: for this relationship can be explained with equal plausibility as evidence of social causation, with the environmental pressures associated with low social status causing psychopathology; or, by contrast, it can be explained as evidence of social selection with pre-existing psychiatric disorder leading to low social status. The latter interpretation is compatible with the position that genetic factors are more important than social environmental factors in etiology.

The problem of finding a basis for determining the relative importance of social environmental and genetic factors in the aetiology of psychological disorder has proved persistent. Neither environmentally oriented epidemiological studies nor genetically oriented twin studies have resolved this issue, in part because previous investigators have not confronted each others' ideas directly in the designs of their research (4, chapters 2 and 3). Studies of social mobility designed to determine the relative importance of heredity and environment in producing the high rate of disorder consistently found in the lower class have also remained inconclusive because of problems of measuring mobility and insufficient data on family histories of psychological disorders (4, chapter 4).

While it is possible to conceive of straightforward approaches to the problem, such as experimental manipulation of presumed pathogenic factors, or multigeneration prospective studies, critical ethical and practical obstacles stand in the way of their execution. Accordingly, an alternative approach to identifying a crucial test of the aetiological issue has been developed. The strategy, more completely described elsewhere (4, chapters 5 and 11), takes the form of a quasi-experiment based on processes of ethnic group assimilation in open-class societies such as our own.

The theoretical foundation of this strategy rests on three assumptions:

(1) That there is an almost universally shared norm in open-class societies that upward social mobility is desirable.

(2) That serious psychological disorder involves disability that decreases the probability of upward social mobility and increases the probability of downward social mobility.

(3) That there is greater downward social pressure on members of disadvantaged ethnic groups than on their social class counterparts in more advantaged ethnic groups.

On the basis of these assumptions, it is possible to derive from opposing social environmental and genetic theoretical orientations alternative predictions about rates of psychological disorder in different ethnic groups *with social class controlled.*

Thus, if the rate of psychological disorder in a particular social class is a function of the strength of the social pressures experienced by members of that class, we should find higher rates of disorder among members of disadvantaged groups. That is, the greater social pressure exerted on these relatively disadvantaged groups, e.g., Negroes and Puerto Ricans in New York City, would be expected to produce an increment in psychopathology over and above that produced by the lesser social pressure, at any particular class level, on members of more advantaged ethnic groups, e.g., white Anglo-Saxon Protestants and Jews in New York City.

By contrast, from a genetic point of view we would expect just the opposite. For if psychological disorder is mainly an outcome of genetic endowment then we would expect the rate in a given class to be a function of social selection processes, whereby the able tend to rise or maintain high status and the disabled to drift down from high status or fail to rise out of low status. Since the downward social pressure is greater on disadvantaged ethnic groups such as Negroes and Puerto Ricans, we would expect more of their healthier members to be kept low in status, thereby diluting the rate of disorder. In contrast, with less pressure to block them, the tendency of healthy members of more advantaged ethnic groups to rise would leave a 'residue' (9) of disabled persons among the lower-class members of these advantaged ethnic groups, thereby inflating the rate of disorder. Thus social selection should function to give a lower rate of disorder in disadvantaged ethnic groups than in advantaged ethnic groups, social class held constant.

Here, then, is a major substantive issue that could turn on what deceptively appear to be simple questions of fact: for example, what are the rates of psychological disorder among Negroes and Puerto Ricans relative to the rates for their class counterparts in more advantaged ethnic groups in New York City? We expect that the answers will be different for different types of disorder and, indeed, for different types of symptomatology even within what, on its face, appears to be a similar type of disorder.

Assuming that in the present study we have the data with which to develop the appropriate measures of psychopathology, a further basic requirement of our quasi-experimental strategy for investigating the aetiological issue involves its replication in diverse research settings and with different sets of advantaged *v.* disadvantaged ethnic groups. For example, Negroes and Puerto Ricans *v.* other ethnic groups in New York City; Indians and Pakistanis *v.* English in London; southern Italians in northern Italian cities; and so on. Only in this way can we hope to demonstrate the generality of the processes of social causation and/or social selection for different types of disorder and different types of symptomatology; for only through such a series of replications can we hope to rule out idiosyncratic genetic and/or social environmental factors that, one could argue, had affected the results in a single setting.

The first test in this series would be in the New York Metropolitan Region, and it is for this test that the present project in Washington Heights is a developmental stage. We specify the Metropolitan Region as a whole rather than Washington Heights for our first major test of the aetiological issue because the quasi-experimental strategy depends upon ethnic assimilation processes that operate at full strength only in the Region as a whole. For example, we know that younger white middle-class couples with children have tended to move out of Washington Heights to more prosperous suburban areas in the Metropolitan Region. This is a part of the assimilation process that we miss, therefore, if we limit ourselves to Washington Heights. Thus, while the diversity of ethnic groups in Washington Heights, a relatively small geographic area, makes it ideal as a place for pilot studies of our problem, especially its methodological aspects, Washington Heights is less ideal for a major test of the aetiological issue employing our quasi-experimental strategy based on ethnic assimilation.

The test in the Metropolitan Region as a whole, and the long range plan for a series of replications, however, has some compelling procedural consequences. In order to control social class across five or six different ethnic groups in a setting such as the Metropolitan Region, a fairly large sample of respondents is required. Given our interest in situational specificity *v.* persistence of symptomatology (4), the measures should be capable of administration at least at two points in time. This argues for measuring instruments that can be quickly and efficiently administered. Consider in this context two sets of findings from the present methodological study.

The first of these is the tendency of community respondents to report more symptoms to psychiatrists than to lay interviewers. We assume, therefore, that the interviews by psychiatrists provide the more complete and accurate descriptions. Yet efficiency would argue that we use lay interviewers in large-scale tests since, as was noted earlier, the field operation in the present study was both expensive and took two years and four months to complete. The two goals, therefore, of efficient procedures on the one hand and accuracy on the other appear to be in conflict.

The second set of findings of immediately compelling interest in this connection centre on SIS *v.* PSS differences according to the two main independent variables of interest in our strategy—class and ethnicity. If the two instruments give different portrayals of the psychiatric conditions of respondents from different class and ethnic groups, obviously we must choose between them—or, more accurately, since each may do some things better than the other, look to each instrument for what it does best.

These considerations argue that, in our future analyses of the data from the present study, we include two additional aims: first, an attempt to construct a composite interview instrument that demonstrably maximizes the advantages and minimizes the disadvantages of those used so far (and, where both are found to fail in a particular area, construct new measures); and second, to test the degree to which the new instrument can be administered by lay interviewers.

This latter aim will mean structuring the open-ended material in both the SIS and the PSS and in so far as possible, making the judgemental scoring system of the PSS an objective scoring system. When we recall the interviewer–reviewer differences on the judgemental measures, such an attempt to make the scoring procedures more objective can be seen to have perils of its own. The reason is that much of the information that was apparently important to the psychiatrists' judgements in the actual interview situation appears to have remained unrecorded in the present scoring systems. A still more objective scoring system must face the problem of possibly eliminating still more important cues.

It may be that our tape-recordings of the interview, included as a luxury in the present study, will prove in the long run to have been necessary to the realization of these aims. For it is the tapes that provide the nearest thing we have to a bridge between our present scoring and recording systems and what went on in the actual interview situation. And upon such a bridge may depend our chances of developing measures that are not only replicable on an efficient basis, but valid as well.

REFERENCES

1. Crandell, D. L., and Dohrenwend, B. P. (1967). 'Some relations among psychiatric symptoms, organic illness, and social class', *Amer. J. Psychiat.* **123**, 1527–38.

2. Dahlstrom, W. G., and Welsh, G. S. (1960). *An MMPI Handbook* (Minneapolis: University of Minnesota Press).

3. Dohrenwend, B. P., (1966). 'Social status and psychological disorder: An issue of substance and an issue of method', *American Sociological Review*, **31**, 14–34.

4. —— and Dohrenwend, B. S .(1969). *Social Status and Psychological Disorder: A Causal Inquiry* (New York: J. Wiley).

5. —— and Chin-Shong, E. (1967). 'Social status and attitudes toward psychological disorder: The problem of tolerance of deviance', *American Sociological Review*, **32**, 417–33.

6. Edwards, A. L. (1957). *The Social Desirability Variable in Personality Assessment and Research* (New York: Dryden Press).

7. Elinson, J., Haberman, P., and Gell, C. (1968). *Community Fact Book for Washington Heights, New York City, 1965–66 and 1960–61* (New York: Columbia University School of Public Health and Administrative Medicine).

8. Faris, R. E. L., and Dunham, H. W. (1939). *Mental Disorders in Urban Areas* (Chicago: Chicago University Press); and, for example, Meyerson, A. (1940). Review of 'Mental disorders in urban areas', *Amer. J. Psychiat.* **96**, 995–7.

9. Gruenberg, E. M. (1961). Comments on Dunham, in *Causes of Mental Disorders: A Review of Epidemiological Knowledge, 1959,* p. 269 (New York: Milbank Memorial Fund).

10. Haase, W. (1964). 'The role of socioeconomic class in examiner bias', *Mental Health of the Poor,* eds. Riessman, F., Cohen, J., and Pearl, A. (New York: Free Press).

11. Jackson, D. N., and Messick, S. (1962). 'Response styles on the MMPI: Comparison of clinical and normal samples', *Journal of Abnormal and Social Psychology,* **65**, 285–99.

12. Langner, T. S. (1962). 'A twenty-two item screening score of psychiatric symptoms indicating impairment', *Journal of Health and Human Behavior,* **3**, 269–76.

13. Leighton, D. C., Harding, J. S., Macklin, D. B., Macmillan, A. M., and Leighton, A. H. (1963). *The Character of Danger* (New York: Basic Books).

14. Manis, J. G., Brawer, M. J., Hunt, C. L., and Kercher, L. C. (1964). 'Estimating the prevalence of mental illness', *American Sociological Review,* **29**, 84–9.

15. Phillips, D. L. (1966). 'The "true prevalence" of mental illness in a New England State', *Community Mental Health Journal,* **2**, 35–40.

16. Spitzer, R. L., Endicott, J., and Fleiss, J. L. (1967). 'Instruments and recording forms for evaluating psychiatric status and history: rationale, method of development and description', *Comprehensive Psychiatry,* **8**, 321–43.

17. Srole, L., Langner, T. S., Michael, S. T., Opler, M. K., and Rennie, T. A. C. (1962). *Mental Health in the Metropolis* (New York: McGraw-Hill).

18. Star, S. A. (1950). 'The screening of psychoneurotics in the army: technical development of tests', *Measurement and Prediction,* pp. 486–547, eds. Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., and Clausen, J. A. (Princeton, N.J.: Princeton University Press).

19. —— (1955). *The Public's Idea About Mental Illness* (National Opinion Research Center, University of Chicago, mimeographed).

Discussion

I

K. RAWNSLEY
Professor of Psychological Medicine
Welsh National School of Medicine
University of Wales

A large number of observations upon the differential prevalence of psychiatric disorders by social class and by ethnic group have been published. The majority of these have employed 'treated' cases as numerators and are not therefore our concern today. Dohrenwend and Dohrenwend have reviewed 43 community studies where untreated as well as treated cases have been ascertained and many of these enquiries have correlated morbidity rate with social class and/or ethnicity.

All three major variables—psychiatric disorder, social class, and ethnic group—raise formidable problems of definition and of precise measurement. In the current paper, Dohrenwend and his colleagues seek to examine the relationship between different measures of psychiatric morbidity when these are applied to groups of known psychiatric patients, to community leaders and to a stratified random sample of Washington Heights residents in New York City.

The study draws freely upon the methods and techniques of two important American researches: the Stirling County study and the Midtown, Manhattan, study. The Dohrenwend inquiry, however, appears to be far less interested than either of its predecessors in developing a theoretical framework linking the development of psychiatric disorder with the social environment.

The paper is by way of an interim report and the analysis of material is still under way. However, important findings have already emerged which touch upon fundamental issues of morbidity assessment. By virtue of the nature and size of the samples studied, psychiatric ailments detected are likely to be, in the main, non-psychotic. No one who is familiar with the effects of observer bias in the assessment of 'hard' morbidity data such as chest X-rays and electrocardiograms will be surprised to learn that the evaluation of a 'Psychiatric Status Schedule' varies considerably depending upon whether one has seen the respondent directly in his social context or not.

Thus, in the current study, judgemental evaluations based on the SIS and PSS show an inverse relationship between 'caseness' or 'impairment' and social class. If the PSS schedules are rated 'blind' by other psychiatrists, the gradient disappears or may even be reversed. It is postulated that the actual interviewers may see more evidence of positive functioning than the reviewers and may therefore discount the importance of some of the symptoms. This may be linked differentially with social class.

'Caseness' or 'impairment' represent Stirling-type ratings or Midtown-type ratings respectively. The judge reviews the whole schedule and makes his assessment. Objective measures based on statistical manipulation of symptom and other data are also being developed in the course of the analysis. Acknowledgement is made, however, to the problem of 'subcultural differences in modes of expressing distress' which might complicate comparisons between social classes or ethnic groups.

This study is endeavouring to tackle, therefore, in a bold and sophisticated way, some of the central problems of morbidity assessment. Preoccupation with observer error and bias, with the importance of acquiescence and social desirability 'sets' and with the effect of these upon inter-class and inter-ethnic prevalence findings is both refreshing and sobering.

The question of validity is, of course, discussed. Both criterion-oriented and construct validity are employed in analysis. An important issue arises here in the matter of the selection of criterion groups. In many previous studies, and to a degree in the present one, patients attending clinics or hospitals are used as a criterion. The whole issue is bedevilled and complicated by the unknown effects of recognition and of formal adoption of the sick role upon the profiles of response to various case-finding instruments. In this respect, therefore, declared cases, i.e., patients who have come up for diagnosis and treatment, may be a rather inappropriate criterion group.

An important feature of the current study is the way in which it seeks to examine the degree of congruence of various instruments in comparing prevalence between categories of individual. Although each individual receives an APA diagnostic category, the main emphasis thus far in the analysis is on degree of 'caseness' or of 'impairment'. Although the authors must be well aware of it, this creates many problems for the comparison of morbidity across social classes or ethnic groups. Faris and Dunham in their classic study found a decided difference in the ecological patterning of schizophrenia as against manic-depressive illness. The same divergences may be found in nosological categories outside the psychotic rubric. They may even be such as to cancel out each other when total morbidity is considered.

One therefore regards the second major focus of this study with certain misgivings and reservations. This is to consider 'the problem of the relative importance of social causation versus social selection factors in social class differences in rates of psychopathology'. For purposes of argument it is rather assumed that the current study will mirror the findings of previous inquiries which have shown a concentration of psychiatric disorder among lower social classes. As was mentioned above, this was found not to be so on the PSS judgemental evaluation made 'blind' by reviewers. However, assuming that the measures in general support previous findings, an ingenious quasi-experimental strategy has been devised using both inter-class and inter-ethnic differential prevalence which, it is thought, will help to settle the causation *v.* selection hypotheses.

Whilst admiring the subtlety of this approach, I must confess to a feeling that the project begins to over-reach itself a little at this point. The tentative theories which are advanced to predict the relative influence of genetic and socio-environmental factors upon prevalence of psychiatric disorder by social class in different ethnic groups have a naïve flavour and beg many questions. One sees the early scaffolding of a 'frame of reference' here of a kind which has tended to have a somewhat stultifying effect upon earlier large scale-community surveys.

Although this is only a personal opinion, I should have thought there would have been scope for a great deal more work on the methodology of assessment of psychiatric morbidity arising from the impressive studies reported in this paper. I believe much more needs to be done along these lines before one would be in a position to tackle such a complex problem as that of causation *v.* selection in social class prevalence differences.

The current study represents in my view a significant growth point in the evolution of morbidity measurement in psychiatric community studies. It combines elements of the American and of the European tradition in this field and the fact that some of the findings are, on the face of it, contradictory and difficult to interpret indicates that it is probably tapping the truth.

2

ROBERT J. KLEINER

Temple University, College of Liberal Arts, Philadelphia, U.S.A.

I shall discuss some general impressions and ideas for new research which I have had from reading the report of Dr Dohrenwend and his associates. First, I was particularly impressed by the thoroughness with which the authors pursued their task and the manifest methodological sophistication of the design. My colleagues and I can sympathize with them in their study because we have had to deal with similar problems in our own research.. It can, of course, be very frustrating trying to unravel factors that may account for apparently different or inconsistent findings obtained by different measuring instruments presumably concerned with the same issue. Yet it was somewhat disconcerting to me to find few references to tests of significance in the main body of the report and I was led to assume either that few fests were made or that the results were not statistically significant. Having made this assumption (which would be contrary to the authors' interpretation), I was impressed by the *similarity* of findings obtained with the two instruments, as well as the *similarities* between the subgroups, and I consequently felt that the detailed discussions of differences which appeared small and insignificant were somewhat belaboured and over-extended.

In some respects the in-patient, out-patient, and prison groups differed from the community groups. However, what impressed me most was the fact that the prisoner group appeared, in almost every analysis, to be sicker than the out-patient group, and in some respects sicker than the in-patients. There may be at least three reasons for this. First, since the prisoner group was a known group, the psychiatrists evaluating their data may simply be projecting their image of what imprisoned men are. Second, the measuring instruments may not be sensitive enough to discriminate among the groups in any other way. And lastly, it may be that the prisoners are, in fact, more mentally ill than one or both of these patient groups. Of course, if the last-mentioned possibility is correct, then our attitudes and actions with respect to penal systems must be radically revised. The implications of this finding would be far reaching. We have just gone through a long period, at least in the U.S., of removing psychiatric patients from local prisons. However, they were placed there because the psychiatric hospitals lacked the space and bed capacity to absorb them and not because of misdiagnosis. Further research is clearly required to determine the reliability of these findings.

When I was reading the introduction, I was particularly interested and excited by the emphasis of the authors on the need to study situational and subcultural determinants of the labelling process or the process of identifying a 'case'. I had hoped that the report would deal more fully with these issues and was somewhat disappointed when they were held out only as issues for future study. I look forward to the authors' discussion of these issues at a later date because their data certainly lends itself to such analyses. Concern with these issues was of particular importance in our own work with data relating social structural variables to rates of mental illness in an urban Negro population. For each of the traditional status measures of education, occupation, and income, there were minimal differences with respect to the determinants of the labelling process studied. On the other hand, the data indicated that the labelling process would lead to under-estimating rate differences between downward mobile and non-mobile groups, as well as between natives and migrants.

In their discussion of 'stressor' factors, the authors paid too little attention to the problem of whether, or under what conditions, factors which are *assumed* to be stress-provoking, e.g., having a disabled father, deceased mother, frequent separations from the family, etc., are really so. In addition, the summing of the number of such factors, as in the Midtown study, does not necessarily mean that we are increasing the total stress in any simple manner. It is also problematic whether equal weight for each stressor factor can properly be assumed. If the factors are, indeed, stress-provoking, they will vary in their importance for the total mental health of the individual. Their relative importance must be determined before we can have a meaningful predictive 'stress' model of mental illness.

I would like to mention now a number of general ideas for new research that derive from this report and other discussions that have been concerned with these issues. Most studies concerned with improving measuring instruments to determine the prevalence and incidence of mental illness assume the validity of some psychiatric frame of reference. For example, the ultimate judge of a 'case' is the psychiatrist, either through direct evaluation of the person or through the studied evaluation of protocols gathered by others. The measuring instruments consist of items that are considered relevant to mental illness by psychiatrists. Even when we concern ourselves with negative attitudes toward psychiatry and lack of sophistication about psychiatry—factors which presumably would distort the 'validity' of prevalence or incidence studies—the 'degree of distortion' always takes a psychiatric point of view as the 'true' base-line. In studies concerned with evaluating these 'artifacts', evaluators are given hypothetical case-histories to diagnose, which are almost always descriptions of individuals that all psychiatrists would agree are 'cases'. Actually phrases such as 'degree of distortion', 'sophistication about psychiatry', 'attitudes toward psychiatry', etc., are different ways of saying—'if only everyone would internalize the psychiatric frame of reference our job would be easier'. The fact of the matter is that people in general have not internalized this frame of reference and I doubt if they even think in these terms until a case is so obvious that no one would disagree about it at all.

Not only do laymen and professionals disagree with each other on this issue, but the professionals disagree with one another. This has been demonstrated by some of the data in the present report which show that two documents, presumably constructed on the basis of a particular frame of reference, do not agree on how to identify a 'case'. But I am less concerned with this problem and more concerned about the typical situation in any community. It is usually subdivided into many different groups, each with its own definition of mental illness. These groups also differ in the way in which individuals are expected to act toward someone so labelled by the group. People look at the problem from different perspectives than the professional and yet the professional depends on this very community for their referrals, for 'case' counts and for the assessment of prognosis. I propose that, as an *additional* direction of research, we need to study mental illness from perspectives (or reference groups) other than the psychiatric ones. Parallel studies are necessary to find out how different groups (1) define mental illness, (2) what they would do about it, (3) who would they consider seeing, and (4) in what order of priority. If we made such studies we could superimpose the findings obtained from one frame of reference on to the findings obtained from another—for example, in the relationship between social class and schizophrenia. The areas of incongruence would give us greater insight into the problems of epidemiological research and their implications for theory. Where the findings are concordant we could make more stable generalizations.

Finally, I would like to point out that what we consider to be the social antecedents of illness may also vary in their meaning for different social groups (or reference groups). Thus in one social context in which occupation is used by the researcher as a measure of status, an inverse relationship between this measure and rates of illness may occur, but in another social context it may not. We usually direct our attention to the occupation scale, to the problem of attitudes toward psychiatry or to sophistication about psychiatry, etc. Rarely do we consider the possibility that occupation itself may not have any special implications for a population in the specific context. For example, if a group of people concentrate in a very narrow range of occupations, it is unlikely that they would use this factor as a measure of prestige or status. Similar reasoning would apply to education or income. Many of our cross-cultural studies use the same measures of status but come up with different results, whereas others come up with the same results. In the latter case we assume that the dynamics of the different situations are the same. More often, however, we are puzzled by 'inconsistent' data. We must know more about the operative symbols of prestige and status in different social groups and social systems (or reference groups), so that we can more adequately study the relationship of such factors as status to mental illness. In the U.S., many studies use the Social Class Index of Hollingshead and Redlich, developed on a sample from New Haven, Connecticut. When this index was used with samples of Negro adults as well as Negro youth, no systematic correlations were obtained. When a social class index for Negroes was developed with a Negro sample and then applied to a large urban Negro population, all kinds of interesting correlations and relationships were obtained.

5

INCIDENCE AND DURATION OF EPISODES OF MENTAL ILLNESS IN A TOTAL POPULATION

The incidence and duration of episodes of mental illness in a total population

OLLE HAGNELL

The Research Centre, Swedish Social Board, Dalby, Sweden, and
Department of Psychiatry, University of Lund, Lund, Sweden

The annual incidence and duration of episodes of mental disorder were studied in the total population of two Swedish villages of 2,550 inhabitants, who were first examined in 1947 and again ten years later, irrespective of where they then lived. It was possible to re-examine 99 per cent of those living.

All intervening mental illnesses were recorded from personal interviews, key informants including their physicians, and the record of psychiatric services. For more detailed descriptions of data-collecting, population, etc., see Hagnell (1) and Leighton *et al.* (4). For a description and examples of what is meant by a 'case of mental disease, see Hagnell (1, 2).

I work here with the *pattern of mental illness episodes*. In my monograph of 1966, only the first incidences were considered. Here I am concerned with any episode observable during the ten years. All those registered and interviewed in the population in 1947 are included, irrespective of whether they consulted a doctor, treated as well as untreated. In this presentation, I deal primarily with: (1) ages at onset of mental illness and (2) duration of neuroses.

The average populations of 1947 and 1957 were chosen as reference groups for each ten-year age-group (of the 1947 population). Of the 2,550 persons investigated in 1947, 253 died during the ten-year period up to 1 July 1957. The observation period has thus been reduced from 25,000 observation years to 24,208·5.

For every episode of mental illness, data were obtained on diagnosis, degree of impairment, and date on onset and termination. Onset and termination were mostly recorded by month, although sometimes this was impossible and I then had to record the quarter and even in some instances the half-year.

TABLE 5.1. *Incidence of all episodes of mental illness during a ten-year period by first episode and subsequent eqisodes*

		Age at onset									
		0–9	*10–*	*20–*	*30–*	*40–*	*50–*	*60–*	*70–*	*80–*	*All*
Men											
Average population		98	204	195	183	194	153	107	90·5	22·5	
1st episode	N	5	15	24	27	26	14	20	20	16	167
	R	0·51	0·74	1·23	1·48	1·34	0·91	1·87	2·21	7·11	
Subsequent episode	N	0	2	9	22	24	13	11	6	0	87
	R	0·00	0·10	0·46	1·21	1·27	0·85	1·03	0·66	0·00	
Totals:	N	5	17	33	49	50	27	31	26	16	254
	R	0·51	0·83	1·69	2·69	2·58	1·76	2·90	2·87	7·11	
Women											
Average population		104	195	155	153	178	154	121	93	11·5	
1st episode	N	2	22	44	45	61	38	20	19	20	271
	R	0·19	1·13	2·84	2·94	3·43	2·47	1·65	2·04	17·40	
Subsequent episode	N	0	9	31	39	47	39	16	8	1	190
	R	0·00	0·46	2·03	2·55	2·64	2·53	1·32	0·86	0·87	
Totals:	N	2	31	75	84	108	77	36	27	21	461
	R	0·19	1·59	4·84	5·49	6·07	5·00	2·97	2·90	18·27	

N = number of persons. R = rate per hundred per year.

The measurement of both relapses and remissions enables us to dispense with thinking of mental illness as a lifetime characteristic of an individual, and we do not miss—as is otherwise so easy—the acute brief conditions and the fluctuating patterns of such illness. The age of the proband is known at the onset of each episode of illness and so the age-specific incidence of the onset of episodes may be calculated.

The incidence of onsets of episodes of mental illness

The curves of the average annual incidence of onsets of episodes of mental illness (Fig. 5.1) are bimodal in both sexes. The men's first mode is at 30–9 years; the women's is at 40–9 years. There are certain differences between the two sexes. For women, the curve rises strongly up to the 20–9 age; it then remains on the whole constant, except for a small maximum in the 40–9 age.

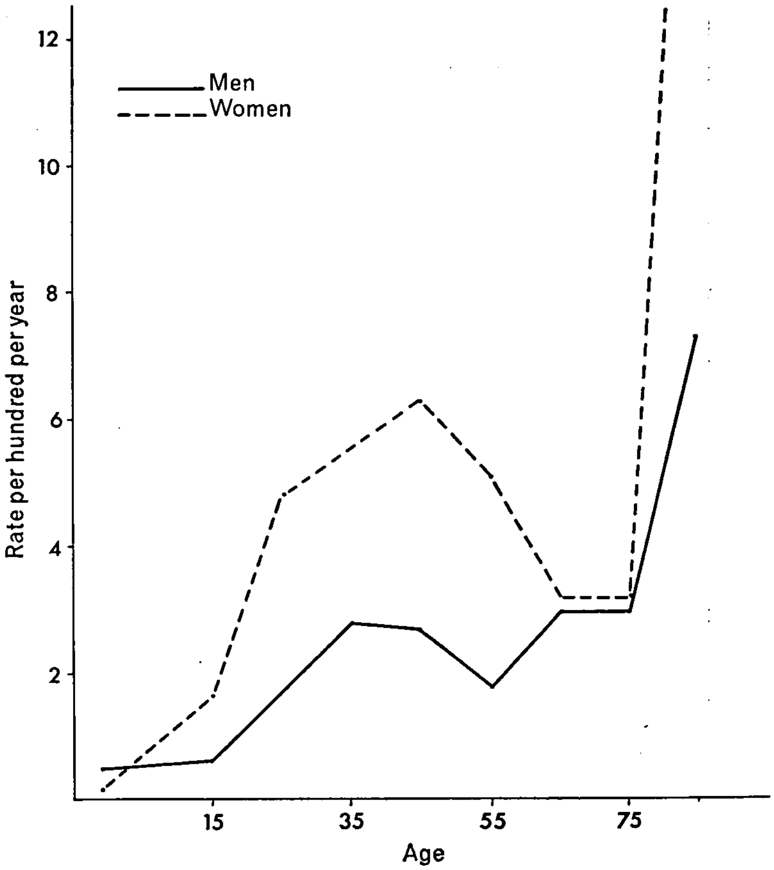


FIG. 5.1. Average annual incidence of all mental illnesses, by sex

From and including the sixties, the average annual incidence falls markedly and then rises again in the highest ages—over 80 years. For men, the curve shows a decline from the thirties to the fifties; however this is by no means so obvious as for women. For women, the first mode is a decade later than for men. The second mode for both sexes is in the highest age-groups, and for 80 and over the ratio is higher still.

Thus it is reasonable to regard the bimodal curves as reflecting two groups of illness: the first group with its onset mostly before the age of 30, the second in the later decades of life. These

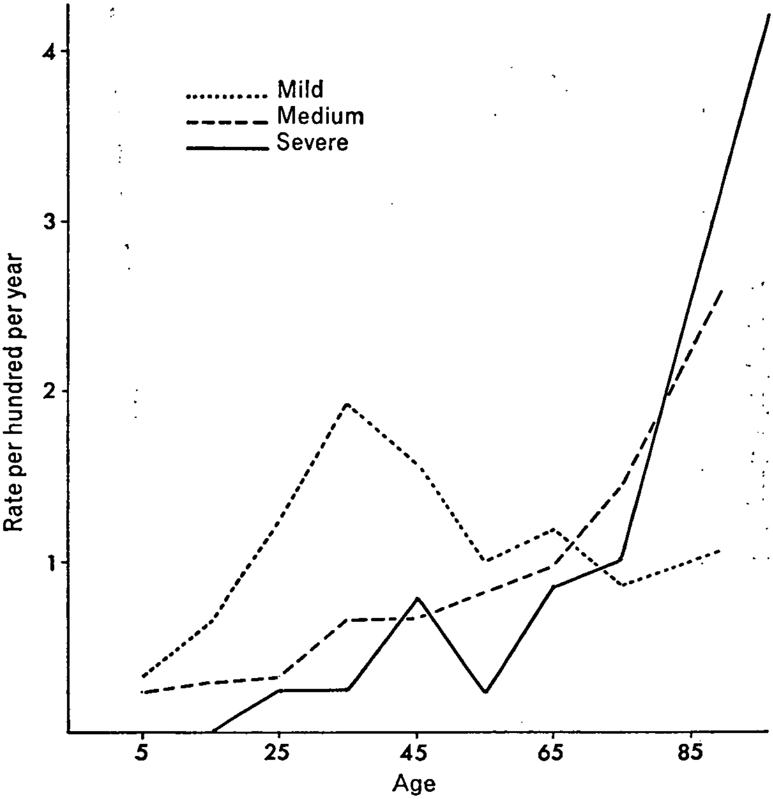


FIG. 5.2. Average annual incidence of onsets of mental illness, by degree of impairment, men

groups may be named the 'middle-years-maximum' (MYM) and the 'rise-with-age' (RWA).

Before continuing with a description of the clinical types, I will discuss two graphs that illustrate the degree of impairment caused by the illnesses. Impairment of the functioning was classed as mild, medium, or severe (for description and examples see Hagnell (1, 2)). *Mild* impairment of functioning is perhaps the most interesting of these grades beause it borders on normality. From Figs. 5.2 and 5.3, we see how the average annual incidence increases strongly in adolescence for women, then remains constant in middle age, decreases in the fifties and continues to decrease with rising age. For men, this is not so pro-

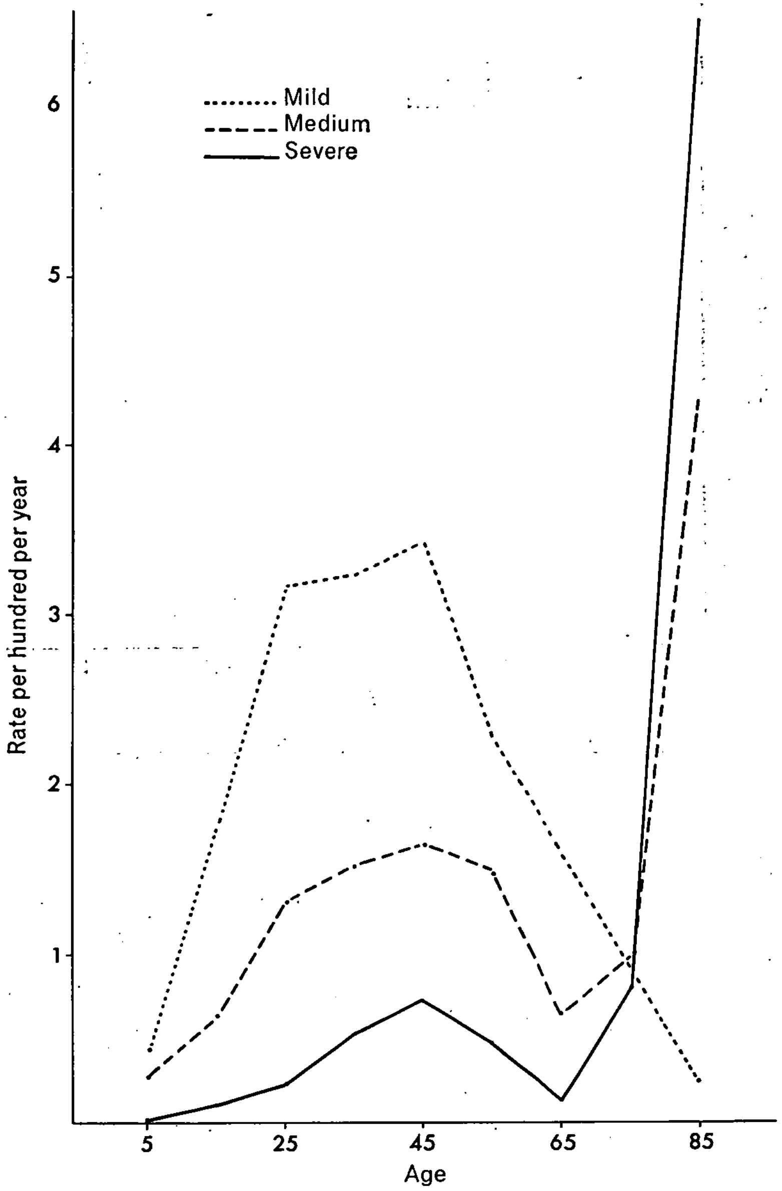


FIG. 5.3. Average annual incidence of onsets of mental illness, by degree of impairment, women

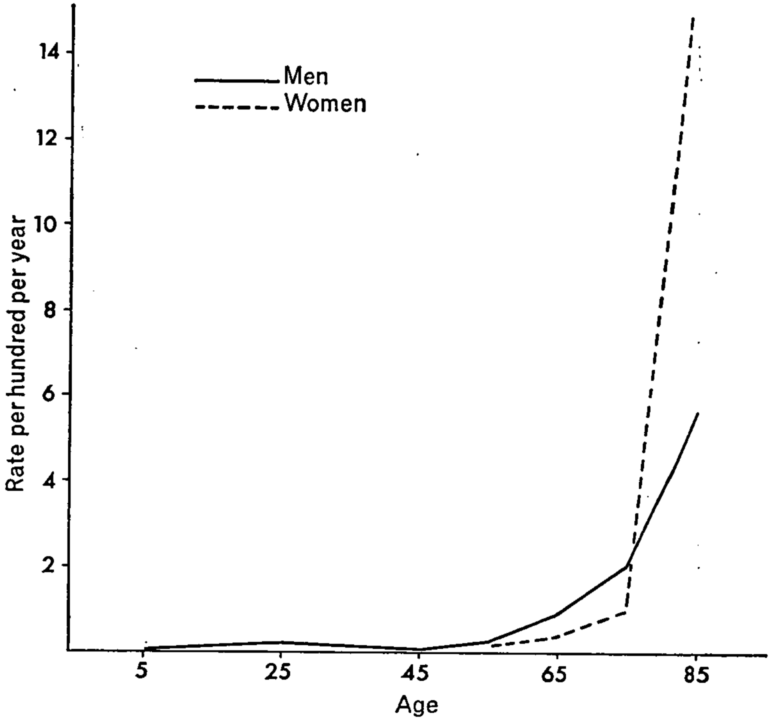


FIG. 5.4. Average annual incidence of brain syndromes, by sex

nounced: there is a marked peak in early middle age and then a gradual decrease. For *medium* and *severe* impairment, the curves for men are almost J-shaped; but for women there are two maxima, one in middle age and the other in the highest age-group.

The 'rise-with-age' (RWA) group of illnesses comprises organic brain syndromes and a combination of organic syndrome with old-age psychosis. In the earlier decades, most of them are post-commotional states. The average annual incidence of brain syndromes begins to rise in men in the sixties but not until the seventies in women (Fig. 5.4). This difference is attributable to the earlier age at which men begin to show the manifestation of cerebral arteriosclerotic brain disease. In a previous report (3), I have shown that the risk of strokes is equal for the two

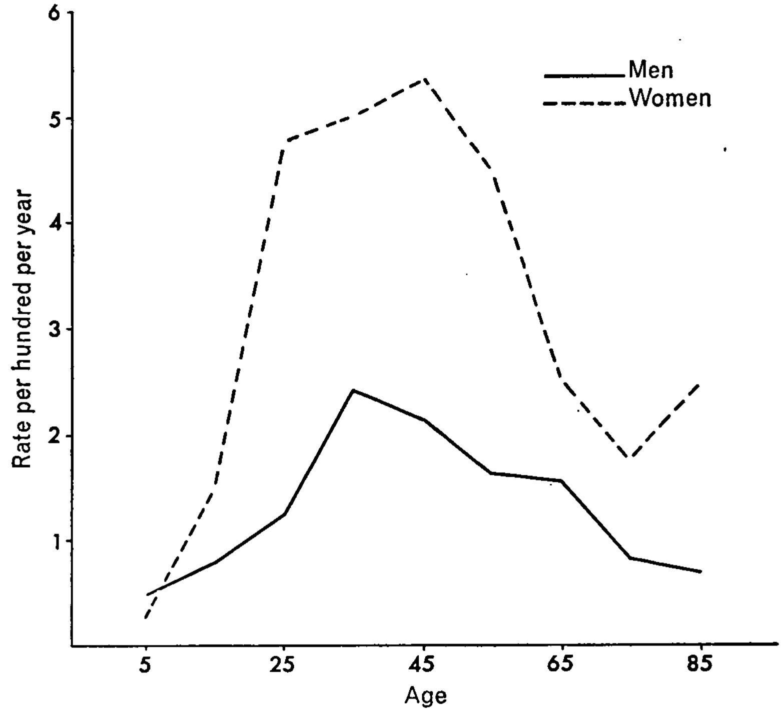


FIG. 5.5. Average annual incidence of neuroses, by sex

sexes but that strokes in men are more often complicated by psychoses in the form of confusions, delusions, etc.

The second group, 'middle years maximum' (MYM), contains mainly neuroses. Neuroses in this study refer to non-psychotic mental illnesses in the absence of obvious brain disease. The age-specific annual incidence rates for neuroses rise rapidly with age from about 15 years to a maximum at age 35 (30–9) in men and 45 (40–9) in women (Fig. 5.5). This supports Shepherd's and Gruenberg's hypothesis that 'neuroses have peaking onsets in the years of early maturity' (5).

The maximum age-specific annual incidence of all neurotic episodes was 5·6 per hundred per year (rate/100/year) in women and 2·4 per hundred per year in men. The age-specific incidence of all neuroses drops sharply with advancing years after reaching these maxima. Some possible explanations of this are:

(1) The illness in older people has a long duration.

(2) Older people attract less attention to their symptoms (perhaps because of reduced impairment).

(3) Somatic diseases or complaints have taken over.

(4) People in these years have better mental health.

An analysis of the onset of episodes of neuroses in these ages shows that most of them have a relatively short duration (Table 5.2). Half of the episodes in men have a duration of less than four months and only two a duration of more than two years. Almost half of those in women have a duration of less than four months; ten (less than 10 per cent) have a duration of more than two years. Thus the lower rate of episodes in higher age-groups cannot be explained by their having longer duration.

TABLE 5.2. *Duration of episodes of neurosis during the ten-year period, showing number of episodes*

	Age at onset				
Duration	*40–4*	*45–9*	*50–4*	*55–9*	*Cumulation*
Men					
1–4 months	10	13	8	5	36
5–8 „	6	6	3	3	54
9–11 „	1	1	1	1	58
12–23 „	2	3		1	64
24 „		1	1		66
Total	19	24	13	10	
Women					
1–4 months	20	15	9	3	47
5–8 „	9	8	7	4	75
9–11 „	4	2			81
12–23	8	5	2		96
24 „	1	2	5	2	106
Total	42	32	23	9	

The second possibility, that older people attract less attention to their symptoms, is not supported by my experience at interviews.

The third possibility, that neurotic complaints are replaced by somatic ones, can to some extent be examined from my data, where somatic illnesses were recorded. The data do not support this, though of course no actual physical examination was made.

The final suggestion, that older people have better mental health, is not improbable. Perhaps, through adaptation, the degree of impairment they experience is so reduced that they can rightly be considered mentally healthy.

The duration of episodes of mental illness

The average duration of episodes of mental illness for psychoses and brain syndromes is about two years, for neurosis eight to nine months (Table 5.3). If the neuroses of old age are excluded, the average duration becomes somewhat shorter, about six to seven months. There is no marked sex difference. The average total length of mental illness per mentally ill person is naturally longer than the individual episodes. The pattern, however, is the same. First episodes last longer than subsequent episodes (note that 'first episodes' are the first episodes during the ten-year period and that they are not necessarily first-ever episodes;

TABLE 5.3. *Number of episodes of mental illness, and mean duration, by sex and diagnosis (the figures in brackets include all those ill at cross-section day, 1 July 1947)*

	Episodes				*Persons*			
Diagnosis	*Number of episodes*		*Mean duration (months)*		*Number of persons*		*Mean duration (months)*	
Men								
Neurosis	188	(198)	8·0	(8·3)	134	(140)	11·2	(11·7)
Psychosis	20	(27)	22·7	(47·9)	17	(24)	26·6	(53·9)
Brain syndrome	46	(54)	28·3	(33·2)	46	(54)	28·3	(33·2)
Women								
Neurosis	407	(430)	8·9	(10·2)	282	(298)	12·8	(14·6)
Psychosis	21	(34)	15·5	(47·4)	19	(32)	17·1	(50·4)
Brain syndrome	33	(43)	31·7	(32·7)	33	(43)	31·7	(32·7)

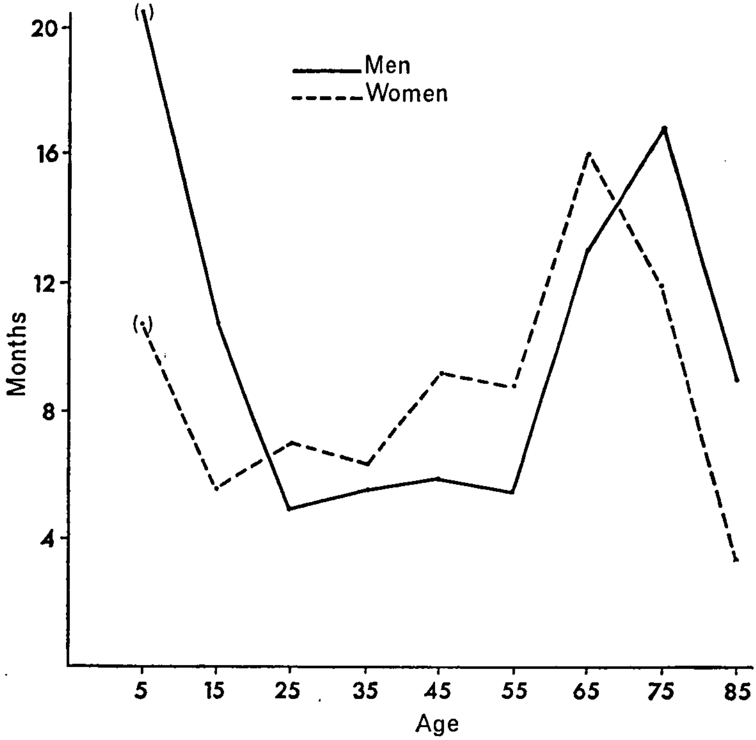


FIG. 5.6. Episode duration of neuroses, by age and sex

but even if this is taken into account, the same pattern still holds good).

Psychoses and brain syndromes are relatively few, and it is therefore difficult to make a more detailed analysis. Neuroses, on the other hand, are common. Episode duration of neuroses by age is shown in Fig. 5.6 and Table 5.4. The duration of episodes is longer in the lowest and the highest ages, but it must be noted that there are relatively few probands in the lowest age-group. Between adolescence and the fifties the mean duration does not vary much. Forty per cent of episodes have a duration of less than three months and 70 per cent of less than six months. Only 4 per cent have a duration of more than three years. Thus there are a large number of neurotic episodes of

TABLE 5.4. *Number of neurosis episodes in the ten-year period (sexes together)*

Age at onset	*Duration of episode (months)*					*Total number of episodes*
	0–3	*4–6*	*7–12*	*13–36*	*36*	
0–9		1	2	4		7
10–19	18	15	8	4	1	46
20–9	50	26	14	10	1	101
30–9	47	46	11	11	2	118
40–9	60	38	27	10	7	142
50–9	43	22	18	14	2	99
60–9	22	15	3	2	8	50
70–9	4	7	5	8	2	26
80–9	1	3		1	1	6
Total	246	173	88	64	24	595
Percentage of total episodes	41	29	15	11	4	100

short duration in this population. It also seemed that many neuroses had a limited course even if untreated.

Summary

The annual incidence and duration of episodes of mental disorder were studied in the total population of a geographical area in Sweden, who were first examined in 1947 and again ten years later, irrespective of where they then lived. The curves of the average annual incidence of onsets of mental illness are bimodal in both sexes. The men's first mode is at 30–9 years (2·7 per hundred per year), the women's at 40–9 years (6·1 per hundred per year). The second mode is in the highest age-groups.

The mean duration does not vary much between adolescence and the fifties: it remains within five to nine months for men, somewhat longer for women. Thus a high proportion of neuroses are of short duration. Many neuroses seem to have a limited course even if untreated.

REFERENCES

1. Hagnell, O. (1966). *A Prospective Study of the Incidence of Mental Disorder. A study based on 24,000 person-years of the incidence of mental disorders in a Swedish population together with an evaluation of the aetiological significance of medical, social, and personality factors* (Scandinavian University Books, Norstedts/Bonniers, Stockholm). (In U.S. Humanities Press, Park Avenue South, New York, N.Y.)
2. —— (1966). *Kasuistik till*, 'A prospective study of mental disorder', mimeo (in Swedish).
3. —— (1969). 'Disease expectancy and incidence of mental illness among the aged', *Acta psychiat. scand.*, Suppl.
4. Leighton, D., Hagnell, O., Leighton, A., Harding, K., Kellert, S., and Danley, R. (1969). 'Psychiatric disorder in a Swedish and a Canadian community: an exploratory study', *Social Sciences and Medicine* (in press).
5. Shepherd, M., and Gruenberg, E. M. (1957). 'The age for neuroses', *Milbank mem. Fd Quart.* **35**, 258.

Discussion

I

H. WARREN DUNHAM

Department of Sociology and Department of Psychiatry,
Wayne State University, Detroit, Michigan, U.S.A.

In his paper Dr Hagnell provides us with some new data on the number of episodes of mental illness and their duration, with a particular emphasis on the neuroses, but based on the data already collected in his monumental survey of mental disorders in two Swedish parishes in 1957. This was the same population that had previously been studied by Professor Essen-Möller in 1947.

With a thoroughness that we have come to expect from Scandinavian epidemiological studies of mental disease, Dr Hagnell has literally showered us with a wealth of statistics, so voluminous in both depth and breadth, that it becomes difficult to pull out those solid facts that will advance our epidemiological knowledge of mental illness. It might therefore be helpful to examine the objectives of such epidemiological studies so as to try and determine how far Dr Hagnell has furthered these objectives.

Objectives central to the epidemiological study of mental illness are concerned mainly with the attempt to develop various statistical measures which will tell us the true incidence and prevalence of mental illness or of a specific mental illness in a given population of risk. Such measures, assuming that exact ones can be developed, are meaningless unless they are compared with similar measures constructed in the same way from other societies. In other words, we expect these measures to lead to a comparative study of the incidence and prevalence of mental disease in the different societies throughout the world. Not only is the epidemiologist concerned with such comparative studies with other societies, but also he is concerned with using these measures to determine if there is a variation in the incidence and prevalence in different parts of the same society.

The two basic needs here, as all epidemiologists working in this area are aware, are (1) to develop universal and valid criteria for determining a case of mental illness and (2) to develop some uniform method for determining a point in time when a given mental illness begins. It is obvious that if no uniformity can be secured with respect to these matters, then the comparative epidemiological study of mental illness is a futile undertaking. This latter statement is made, of course, under the assumption that specific types of mental illness will have a uniform character throughout the world wherever human beings are to be found. However, if one makes the opposite assumption that certain types of mental illness are relative to a given cultural setting, then it is only necessary to develop means for meeting these two needs within a given society, with no thought to the possibility of using these measures in other societies.

Of course, in the development of a carefully constructed incidence measure for a given society, as in Dr Hagnell's study, the over-all objective is to provide clues to the nature and function of specific aetiological factors. To this end it is often desirable to determine the extent to which incidence measures are related to selective demographic, social, and cultural variables. Such correlations may suggest where to look further in a given community for the operation of a significant but unknown aetiological factor, for the most compelling objective of epidemiological study of mental illness is to derive hypotheses that may be tested by further epidemiological study or by the use of other methods. In addition, of course, to the search for aetiological factors via an incidence or prevalence measure of mental illness for a given society, such a measure may be useful in the planning of therapeutic programmes, the organization of mental health facilities, and the development of preventive programs.

Dr Hagnell, as reported in his monograph, has indeed attempted to develop incidence measures and he has utilized for his population of risk that proportion of the population which is free of mental disorder at a given point in time but which will subsequently develop a mental disorder. In the present paper, however, Dr Hagnell does not make clear which criteria for determining mental illness he has used in selecting the cases upon which his percentages are constructed. Thus, even though in this paper he is primarily concerned with counting episodes of mental illness and measuring their duration, he should have spelled out clearly the criteria that he has used in counting cases. Considering the various psychiatric orientations throughout the world, we must not forget the extent to which cultural values and preferences may go in the determination of a case of mental illness. Perhaps it might be more desirable in future epidemiological studies to depend primarily on such operational definitions of a case as going to a psychiatrist, entering an out-patient clinic or entering a mental hospital. The definition of a case by the 'author's diagnosis', as used by Dr Hagnell in his study, may present more difficulties than utilizing the social process as the selective agent.

In the present paper Dr Hagnell is first concerned with charting the epidemiology of neurosis; and secondly he is concerned with presenting various statistics on the number and duration of episodes for neurosis and psychosis as a basis for the development of new hypotheses. While he takes great pains to develop statistics that describe the number of episodes and their duration, certain questions can be raised, on the basis of the nature of available data, which challenge the value of such indexes. For example, can we say that the duration of a neurotic or a psychotic episode can be measured with any precision? Considering the fact that there is often a difficulty in determining when a given episode of the disease begins, how can one measure its duration from the beginning to the end? Here, the attempt seems to be to put the neurosis or psychosis into the model of a physical illness where one can chart with greater precision the beginning and the end of the illness. Then, too, it would seem that Dr Hagnell has taken on an almost impossible task when he states that termination of an episode is measured by recovery, by death, or by the end of the study. It is obvious here that in those cases where termination was measured by either death or the end of the study, duration of illness is a false measure. A further point that may produce raised eyebrows is Dr Hagnell's attempt to distinguish between the duration of an episode and the length of a mental illness, which he calls an unknown characteristic.

It is impossible in this short discussion to deal adequately with the numerous statistical data which Dr Hagnell has amassed. However, a few items may be worth attention. Thus, Dr Hagnell reports that the age-specific annual rate for neuroses in women seems to be approximately twice that for men, while at the same time the duration of neurosis is about the same for men and women. His finding that the average duration of a neurotic episode is equal for the sexes, while the average duration of psychosis is seven months longer for men than for women, provides grounds for much speculation. It is difficult to accept the validity of these latter figures. Certainly the social and therapeutic pressures to achieve a remission are as great and perhaps greater for men as compared to women. While this is so for total psychoses, it is not so for schizophrenia, where the average duration of episodes is seven months longer for women than for men.

I have tried in these remarks to examine Dr Hagnell's findings in relation to the major objectives of epidemiological study and to point to some findings that are grounded in certain unresolved issues. His analysis, while it provides us with new data, refines the data in a way which cannot be justified with the tools currently at our disposal.

2

E. H. HARE

Bethlem Royal and the Maudsley Hospitals, London

The Lund survey is unique in making use of a prospective method to ascertain incidence and represents an important advance in technique. But the techniques of psychiatric epidemiology, because of the intractable nature of its material, are still relatively primitive, and our chief concern at this stage of its development must be as much to improve technique as to find out factors or disprove hypotheses. In my discussion, therefore, I will consider certain aspects of the Lund study, as reflected in Professor Hagnell's present paper, from the point of view of their technical strength.

First of all, I think we should notice the term 'total population'. The term is used here in the sense of a geographically defined population, but with the implication (I rather feel) that such a population is a 'real' one and that other populations, in the epidemiological sense, are somewhat artificial. I would not take that view myself. Other studies which have had prospective value have been based on quite different populations: the Cambridge-Somerville youth study (11), for example, was based on schoolchildren, and the British National Survey of Health and Development (2)—which carries the potential for psychiatric information of the utmost value—was based on all children born in Britain during a particular week in 1946. Any defined population is a proper subject of study by the epidemiological method, and a geographical population is not necessarily the most suitable even for incidence studies. It may not be easy to determine whether such a population is representative of the population at large and, indeed, Professor Dunham found in his Detroit study (3) that the mobility of people in towns may be so high that the regular inhabitants are out-numbered by the birds-of-passage.

This leads on to the question of what precisely is meant by the *incidence* of mental illness in a population. Quite a high proportion of the population of Lund (more than 20 per cent of those who remained alive) had moved from the district in the ten years after 1947. These migrants were traced and their illness recorded. But can we be sure that they were exposed to the same stresses as those who remained behind or that they were a group with the same constitutional risk as the stay-at-homes? They certainly differed considerably in age distribution, and probably also in marital status and other characteristics. It would be interesting to know if the incidence of mental illness in the migrants was in fact different from the rest. When one makes a longitudinal study the population one starts with becomes split up in various ways, as a result of the interaction between secular events and individual constitution, and thus becomes exposed to environments which probably differ in their influence on mental health. This indeed was the trouble experienced by Wilner and his team in Baltimore (14). They attempted a longitudinal comparison of the health of families who moved to a new housing estate with those who remained in the old slum but they found that a very high proportion of those who were supposed to have stayed behind had in fact moved away in a short time. There is the further difficulty that even if one confines attention to the non-migrants, the longer one studies them the more the physical and cultural environment will change and with it, to an unknown degree, its effect on mental health. Professor Hagnell has recognized this in saying that even if it were possible to follow a population from the cradle to the grave, we cannot be certain that this would be an ideal method. We might conclude that there is perhaps an optimum period of time for which a population should be studied in order to determine its incidence of mental illness. If the population of Lund were to be studied once again, it might give some answer to this.

Next I should like to consider the problem of diagnosing a mental illness in a non-clinical setting. The Lund study was based on the assumption that a psychiatrist, having requested an interview with a member of the public and having talked with him for say half an hour in his home, is in a position to make a psychiatric diagnosis. Leaving aside those who had in fact sought medical help, we may ask whether a person can be said to have mental *symptoms* if he has never 'complained' of them but only volunteered them in response to direct questions asked at a requested interview. Reliance on symptoms as a means of assessing mental illness has been deplored as humiliating (10), but the use by Hagnell of such diagnostic categories as 'tiredness' does not really overcome this and the difficulty remains that if one has to rely on symptoms (or even on an assessment of the person's general behaviour) one has to consider his threshold of complaint. Everyone has been anxious at times and everyone has been tired, but the degree to which a person expresses his concern about such experiences may depend more on his constitution and sophistication than on their intensity. Kessel and Shepherd (8) studied the health of people who seldom attended a doctor and found they had the same amount of minor illnesses as the frequent attenders but were less worried about their health and scored lower on a test of neuroticism. A further problem arises when (as in the Lund study) a diagnosis implies the presence of mental symptoms which are 'so obvious that we think something should be done therapeutically', because clinicians differ in their views of what can or should be treated. Taylor, in his study of general practitioners (13), concluded not only that 'the better the clinician the less he diagnoses neurosis', but also that 'most neurotic reactions to environmental problems never get beyond the general practitioner, either because he solves them or because he knows them to be insoluble'. Again, minor mental ill-health surely grades imperceptibly into normality and any attempt to draw a dividing line ('a dividing fog' might be a more appropriate term) by a purely subjective impression at interview seems hazardous. Leighton, and Srole and Langner, in their big American surveys, accepted yes-replies in a structured interview as indicators of mental illness. Another approach is represented by that of Kessel and Shepherd (7), who concluded that, at least under the conditions of the British National Health Service, no important psychiatric illness is missed by general practitioners.

We need to consider not only the value of making a psychiatric diagnosis in the setting of a non-clinical interview, but also the reliability of the information obtained. Respondents had to recall episodes of mental ill-health over the previous ten years and not only what these were like but when each began and ended. How far is a person's memory reliable here? Of course, where it can be checked by medical records that is satisfactory but in those cases the interviews would not have been necessary. Cartwright (1) has studied some aspects of this problem in a morbidity survey and it would be of interest to know how far, in the Lund survey, the respondents' memories checked with medical records. We need also to consider the reliability of the interviewer's judgement in making a diagnosis. I do not think this was studied in the Lund surveys. Leighton's team (9) have reported a comparison of a psychiatrist's assessment at interview with an assessment made by a group of psychiatrists from interview schedules obtained by non-medical interviewers and found a gratifying level of agreement. But the reliability of the psychiatric interviewer has not been adequately demonstrated in the field surveys. Dr Shaw and I, in a morbidity survey in Croydon (6), attempted this for the assessment of personality traits and found very poor agreement when two psychiatrists interviewed respondents at an interval of about one month; but we found a much better agreement on information obtained from the use of a structured questionary and this was so whether the interviewers were medical or lay.

In his study Professor Hagnell personally interviewed over 3,000 people, mostly during a single year, which must have meant an average of about eight half-hour interviews every day. By this standard, the size of his population is heroically large. Nevertheless, for statistical purposes it is still on the small side, and for this reason the significance of some of his findings stands on rather thin ice. It is also perhaps worth noting that the duration of a mild neurosis may depend a good deal on the expectation of physicians and relatives as to how long such a condition ought to last; in other words, on how long they feel it reasonable on the whole for a person with such a condition to be given the status of a patient. Thus the duration of the patient's complaints may be conditioned by the 'expectancy therapy' of the local community, and this expectancy may vary in time and place and might, for instance, account for Hagnell's average duration of six months for neurosis in Lund, while Shepherd and Gruenberg (12) suggested between one and two years for the cases they studied in New York.

All these considerations lead me to make two general suggestions. First, the reliability of a psychiatrist's diagnosis of a non-patient in a non-clinical setting has not yet been adequately assessed. Unless it could be shown to be reasonably high—and such evidence as there is suggests it may be very low—the value of this method as a tool for epidemiology must remain in doubt, especially as the expense of the skilled time is likely to allow only a relatively small population to be covered. The information obtained by non-medical interviewers using a structured questionary has been shown to be fairly reliable, though of course it is not possible to comment on its validity without a knowledge of the reliability of the psychiatrist's personal interview. Perhaps we ought to think in terms of two types of psychiatric diagnosis: a *clinical* diagnosis, made in a clinical setting by a psychiatrist who is going to assume responsibility for the case, and an *epidemiological* diagnosis, made in the field for purely research purposes. The primary function of a clinical diagnosis is to serve as a general guide, to the particular psychiatrist who makes it, in his management of a particular patient's case; reliability is a secondary consideration. But for an epidemiological diagnosis, reliability is of the essence because the aim is to provide data which can be generalized.

Secondly, the problem of where to draw the line between normality and minor degrees of mental disorder is insoluble in the context of a single population, but in a comparison of the mental health of two similar populations the problem can be resolved, for if the same technique of case-finding is used for each population it does not matter greatly where the line is drawn.

REFERENCES

1. Cartwright, A. (1963). 'Memory errors in a morbidity survey', *Milbank mem. Fd Quart.* **41**, 5–24.

2. Douglas, J. W. B. (1964). *The Home and the School* (London: MacGibbon and Kee).
3. Dunham, H. W. (1965). *Community and Schizophrenia: An Epidemiological Analysis* (Detroit: Wayne State University Press).
4. Essen-Möller, E. (1956). 'Individual traits and morbidity in a Swedish rural population', *Acta psychiat. Scand.*, Suppl. 100.
5. Faris, R. E. L., and Dunham, H. W. (1939). *Mental Disorders in Urban Areas* (Chicago: Hafner).
6. Hare, E. H., and Shaw, G. K. (1965). *Mental Health on a New Housing Estate* (London: Oxford University Press.)
7. Kessell, N., and Shepherd, M. (1962). 'Neurosis in hospital and in general practice', *J. ment. Sci.* **108**, 159–66.
8. —— and Shepherd, M. (1965). 'The health and attitudes of people who seldom consult a doctor', *Med. Care*, **3**, 6–10.
9. Leighton, A. H., Leighton, D. C., and Danley, R. A. (1966). 'Validity in mental health surveys', *Canad. psychiat. Assoc. J.* **11**, 167–78.
10. Lewis, A. J. (1961). *Comparative Epidemiology of the Mental Disorders*, eds. Hoch, P. H., and Zubin, J., chap. 11 (New York: Grune and Stratton).
11. McCord, W., Porta, J., and McCord, J. (1962). 'The familial genesis of psychoses', *Psychiatry*, **25**, 60–71.
12. Shepherd, M. and Gruenberg, E. M. (1957). 'The age for neuroses', *Milbank mem. Fd Quart.* **35**, 258–65.
13. Taylor, S. (1954). *Good General Practice* (London: Oxford University Press).
14. Wilner, D. M., Walkley, R. P., Schram, J. M., Pinkerton, T. C. and Tayback, M. (1962). *The Housing Environments and Family Life* (Baltimore: Johns Hopkins Press).

6

EVALUATION OF MENTAL HEALTH SERVICES

Evaluation of mental health services

ELMER A. GARDNER
Professor of Psychiatry and Director
Community Mental Health Center, Department of Psychiatry
Temple University, Philadelphia, Pennsylvania, U.S.A.

My association with the Monroe County Case Register (3), and my more recent involvement with the development of a community mental health centre in an urban ghetto has led me to think a good deal about the methods and uses of program evaluation and the evaluation of mental health services.

Edward Suchman (17) distinguished between 'evaluation' and 'evaluative research'; *evaluation* refers to the 'social process of making judgments of worth' while *evaluative research* might be restricted to the utilization of scientific research methods and techniques for the purpose of making an evaluation. Evaluative research increases the possibility for proving rather than asserting the worth of some social activity. Suchman further noted that despite notable exceptions 'most of what passes for evaluative research in most fields of public service, such as health, social work and education, is very poor indeed'; and Carstairs (2) in discussing evaluative research has stated that 'objective evaluations of the effectiveness of new procedures have seldom, if ever, preceded their gaining currency in psychiatric practice' and, further, that 'in the practical politics of psychiatry, just as in the reaching of political decisions in government, persuasion has been the major factor'.

Given this state of affairs, how can we increase the utility of evaluative research so that we can prove rather than simply assert the worth of our programs? Obviously, methodological progress and a constant reluctance to compromise experimental design will enhance our product as a force for effecting programme change. But I doubt that this will be sufficient to assure utilization of findings from evaluative research. Rather, the utility and subsequent implementation of these findings will depend, in large measure, on their incorporation into a service system committed to planned change. In the absence of such a system, even studies which have had few methodological flaws and which have provided a wealth of relevant data, are not widely known and have not been generally translated into actual program operation. The studies of Pasamanick *et al.* (14) and of Sainsbury and Grad (5, 6) are cases in point.

The study of schizophrenic patients by Pasamanick *et al.* (14), one of the best attempts to utilize experimental design in the evaluation of a service pattern, correlated the type of treatment intervention with outcome criteria. The primary questions in this study were:

(1) Whether home care for schizophrenic patients was feasible.

(2) Whether drug therapy was effective in preventing their hospitalization.

(3) Whether home care was, in fact, a better or poorer method of treatment than hospitalization.

To answer these questions, the investigators studied patients with a diagnosis of schizophrenia admitted to a state hospital or referred by a variety of community agencies. The subjects were assigned randomly to one of three groups: home care with psycho-active drugs, home care with placebo medication, and hospitalization (control group). All home care patients were visited regularly by public health nurses and evaluated regularly over a two-year period. Outcome criteria consisted of the frequency and duration of hospitalization, psychiatric and psychological assessment, social functioning, and the extent to which significant others perceived the patient as a 'problem'. The investigators concluded that each of their three primary questions could be answered affirmatively.

All of the groups had improved considerably after six months and showed little or no improvement during the remainder of a two-year follow-up period. Given this comparability of outcome (at least as measured by the study criteria), one might wonder how to make the choice between home care and brief hospitalization (question number 3 of the study). The investigation indicates that home care could be used for a significant proportion of schizophrenic patients, particularly those with families, and thereby reduce the resident patient population in public mental hospitals. This could provide an improved staff/patient ratio, but assumes that staff would remain in the public mental hospitals, willing to care for the residue of sicker patients[1] and that the 'vacancies' produced by preventing hospitalization of these schizophrenics would not be filled by other patient groups. Certainly, as noted by the investigators, much of the movement in community psychiatry and toward community mental health centres had represented a new era of 'moral treatment' (1) and an effort to eliminate *chronic* (not all hospitalization) with the socially debilitating effects of such total institutional care (4, 9). However, community psychiatry and the prevention of hospitalization (usually taken to mean all hospitalization) have become so intertwined in the U.S. that a recent report was issued by the Group for the Advancement of Psychiatry (8) to 'protest against a current bandwagon movement exemplified by such slogans as "Keep patients out of psychiatric hospitals as much as you can. Use the psychiatric ward of a general hospital instead of a mental hospital".'

Another possible advantage of home care over hospitalization could be an economic one, in that home care might be less costly than even brief hospitalization. The investigators note, however, that the cost of such home care programs would depend ultimately on the provision of collateral services and the ability to utilize either non-mental health professionals or non-professionals. This raises a number of questions regarding manpower and the feasibility of utilizing a variety of other kinds of manpower in future mental health programs. Home care might allow for the treatment of larger numbers of patients and could, thereby, represent a response to another aspect of the impetus toward community psychiatry: the growing awareness, through a number of population studies (10, 11, 12, 13, 16), that there is a large number of persons with mental or emotional disorders who need treatment but never receive it. Most of these studies, however, have dealt with non-psychotic disorders where the question is not one of home care versus hospitalization, but home care or some variety of out-patient care versus no care at all.

Whether one would choose home care or brief hospitalization

---

[1] This presumably would pose more of a problem in the U.S.A. than in Britain or other countries with national health services.

might also depend, not only on the acceptance of a patient by the family, but also on some judgement regarding the degree of burden placed on the family or community and the type of influence that acutely symptomatic patients might have on others. This factor, the impact or burden on family and community in the provision of community-based mental health care, was studied by Sainsbury and Grad in their comparison of Graylingwell Hospital's community-based program for Chichester with the hospital-centred services for Salisbury (5, 7). The community-based program for Chichester provides a full range of services, including domiciliary and day-hospital care, while the services for Salisbury have remained limited to in-patient and clinic care.

In general, the Salisbury patients improved slightly more than the Chichester group. The major difference in outcome between the two services occurred with the families of neurotic or depressed patients who were under 65 years of age and admitted to the hospital for a brief time during the two-year study (15). These families suffered particularly with financial problems and complained of mental ill-health in themselves. A comparison of these two patient populations revealed that the Salisbury program provided more active social service with more after-care visits. Thus, the community-based service appeared to do equally well with the more seriously disturbed patients and those patients who presented the most severe burden to their families, but did not seem to be providing adequate social support or after-care to the younger neurotic patients who never burdened their families severely but did continue to cause problems throughout the two-year follow-up period.

One might expect patients with less serious disturbance to be referred earlier in the course of their illness to a community service and that the matched patient samples from the Chichester and Salisbury districts might not be entirely comparable. This was discussed when Sainsbury and Grad presented their findings at a Milbank Memorial Fund conference in 1965 (6).

There is also some reason to question the diagnostic comparability of the patients referred to the two services. The investigators did compare the two referred populations by ranking the frequency with which 28 selected symptoms had been recorded for both populations and found a high rank correlation (0·88), thus implying that the two populations had a similar range of mild and severe symptoms. An examination of the distribution of those symptoms particularly associated with psychosis, namely delusions and hallucinations, revealed that these symptoms were recorded for 28 per cent of the patients treated in the hospital-centred service compared with 15 per cent of the patients treated in the community-based service. Given the comparable diagnostic distribution, this suggests that the psychiatrists in the Salisbury service were utilizing more stringent criteria in making their diagnoses, and that despite the diagnostic similarity and despite the comparability in rank order of symptoms, the population referred to the Salisbury service may, in fact, have been more severely ill than the population referred to the Chichester service. Thus, the better outcome as measured by role performance and the greater relief of family burden in Salisbury would be even more striking if, in fact, this service was treating more severely ill patients. Grad recognized this difficulty in comparing the populations and stated in one paper: 'We would need to see how patients with identical illness fared in both services . . . The groups are not large enough to permit the cross tabulation required for analysis such as employment outcome by age, sex, diagnosis, and length of illness.' She doubted, however, whether the interpretations gleaned from the smaller matched groups would have been any more meaningful than the general trend noted.

The inability to utilize an experimental design in this study and the difficulty in all such studies of controlling many variables make it difficult to judge the relationship between the type of intervention studied and the outcome criteria. For example, the amount of after-care provided by the social service in Salisbury, compared with the lack of such care in Chichester, suggests that the distinction between a 'community'-based and a 'hospital'-based service was not a sharp one. Furthermore, hospitalization occupied only a relatively small proportion of the total observation period for most patients. Thus, other knowledge about care during this period becomes especially significant in evaluating the type of intervention and the relationship between this and outcome at the end of two years, i.e., the amount of after-care, frequency of home visits, and visits to the general practitioner provided by the community-based service.

As I reviewed these studies and other writings on the evaluation of mental health services, a number of other questions have come to mind, particularly about the utility of evaluative research:

(1) How much value is there in the conduct of evaluative studies when these studies are not part of any system of planned change, and when it is unlikely that they will have much, if any, impact on the actual provision of mental health services? I have been struck by the minimal influence that most evaluative research has had on the organization or operation of most mental health services, at least in the U.S. For example, a major program has been launched with NIMH providing massive funding to establish mental health centres throughout the U.S. Yet this funding body, which supports most of the research in the mental health field, has not initiated a systematic evaluation of the centres' program and even an exemplary investigation such as that reported by Pasamanick has had little influence on the amount of home care regularly provided by the centres. Similarly, I wonder how often the 'burden on the family' enters into a choice between in-patient and out-patient care in most mental health programs. Have the findings reported by Sainsbury and Grad produced much change, even in the programs for Chichester and Salisbury?

(2) If the conduct of a proper evaluative study necessitates definition of program objectives, which in turn require explicitness about values and priorities for establishing programs, then who should be responsible for establishing these values or priorities, and who should spell out the program objectives upon which our evaluative studies must rest?

(3) Why have we not promoted the development of systems for funding certain essential evaluative studies, communicating the knowledge from such studies to the responsible mental health professionals and ensuring that such knowledge will effect changes in mental health service delivery patterns?

Gruenberg (9) addressed himself to the issue of determining values, priorities, and objectives in the evaluation of mental health programs. Operational research, he noted, determines or helps to determine whether we realize the desired values, but it does not determine those values. Those who advocate reform must specify the consequences of that reform. Thus, the objectives of the reform must not be seen only from the viewpoint of the investigator who attempts to evaluate the reform. The priorities governing evaluative research should be determined by a larger body or system rather than the individual investigator, and hopefully, by a system which can function with more than short-term goals. Gruenberg documented the objectives advocated by the early reformers in mental hospital care, particularly the objective of reducing the amount of chronic disability associated with chronic mental disorders. This approach to the contemporary evaluation of mental health programs is rarely used and, as noted by Gruenberg, it is even more rare for such an evaluation to precede the growing acceptance of new procedures or services in psychiatric practice.

Finally, I would draw attention to the possibility that those of us who conduct evaluative studies have also resisted a system of planned change. Although we each may recognize the usefulness of evaluative research, we may, nevertheless, be fearful of planned change in mental health services and the possible loss of autonomy which would accompany such a system. If we had sufficient commitment to evaluative research and planned change in mental health programs, we would be working toward a major revamping of the system in which we operate. We would not be content to guess at program objectives or to independently establish the priorities of mental health services. We would not be content to be so limited in our ability to communicate easily and widely and to have most mental health professionals responsible for the provision of services listen mainly to the enthusiastic proponents of some new fad. Without a much more systematic effort, much of what we are doing may be simply game-playing.

As a group, those of us in attendance at this conference could exert considerable influence on local or national funding bodies such as NIMH or the National Health Services, to make evaluative research a vital part of planned change in mental health services. We cannot conduct evaluative research within a system more appropriate for basic research, depending solely upon the scientific validity of our findings. Unless we recognize evaluative study as part of administrative research, methodological progress will have little meaning, and mental health programs will continue to be determined by enthusiasm and salesmanship.

REFERENCES

1. Bockoven, J. S. (1963). *Moral Treatment in American Psychiatry* (New York: Springer).

2. Carstairs, G. M. (1968). 'Problems in evaluative research', in *Community Mental Health*, pp. 44–62, eds. Williams, R. H., and Ozarin, L. D. (New York: Jossey-Bass).

3. Gardner, E. A. (1967). 'The use of a Psychiatric Case Register in the planning and evaluation of a mental health program', in *Psychiatric Research Report 22 (April)* (American Psychiatric Association).

4. Goffman, E. (1961). *Asylums* (New York: Anchor Books).

5. Grad, J. C., and Sainsbury, P. (1963). 'Evaluating a community care service', in *Trends in the Mental Health Services*, eds. Freeman, H., and Farndale, J. (London: Pergamon Press).

6. —— —— (1966). 'Evaluating the community psychiatric service in Chichester: results', *Milbank mem. Fd Quart.* 44, 246–78.

7. —— (1968). *A Two-year Follow-up in Community Mental Health*, eds. Williams, R. H., and Ozarin, L. D. (New York: Jossey-Bass).

8. Group for the Advancement of Psychiatry (1969). 'Crisis in psychiatric hospitalization', *Report No. 72* (New York: Mental Health Materials Center).

9. Gruenberg, E. M. (1967). 'Can the reorganization of psychiatric services prevent some cases of social breakdown?', *Psychiatry in Transition*, ed. Stokes, A. B. (Toronto: University of Toronto Press).

10. Hagnell, O. (1966). *A Prospective Study of the Incidence of Mental Disorder* (Lund: Svenska Bokforlaget).

11. Leighton, D. C., Harding, J. S., Macklin, D. B., Macmillan, A. M., and Leighton, A. H. (1963). *The Character of Danger. The Stirling County Study*, vol. iii (New York: Basic Books).

12. Locke, B. Z., and Gardner, E. A. (1969). 'Psychiatric disorders among the patients of general practitioners and internists', *Publ. Hlth Rep.* (*Wash.*), *84*, no. 2.

13. Pasamanick, B., Roberts, D. W., Lemkau, P. W., and Kreuger, D. B. (1957). 'A survey of mental disease in an urban population: 1. Prevalence by age, sex and severity of impairment', *Amer. J. publ. Hlth*, 47, 923–9.

14. —— Scarpitti, F. R., and Dinitz, S. (1967). *Schizophrenics in the Community* (New York: Appleton Century Croft).

15. SAINSBURY, P. (1968). 'Selecting the patient: the in-patient for discharge', *Proceedings of the Royal Society for the Promotion of Health, Conference on Psychiatric Care in the Community.*
16. SROLE, L., LANGNER, T. S., MICHAEL, S. T., OPLER, M. K., and RENNIE, T. A. C. (1962). *Mental Health in the Metropolis: the Midtown Manhattan Study,* vol. i (New York: McGraw-Hill).
17. SUCHMAN, E. A. (1967). *Evaluative Research: Principles and Practice in Public Service and Social Action Programs* (New York: Russell Sage Foundation).

Discussion

J. K. WING

MRC Social Psychiatry Research Unit
Institute of Psychiatry, University of London

Dr Gardner has asked some important questions. His first deals with the use of evaluative studies which are not part of a system of planned change and which are unlikely to have much impact on the actual provision of services. This question is loaded since one can never be sure in advance how much impact the work will have. But even if one were indeed certain that no immediate improvement in services would result, there are still good reasons for undertaking evaluative research. Investigating the way in which a social environment affects clinical condition should produce knowledge which will be worth while whether it is concerned with the provision of services or not. For example, we undertook some studies of institutionalism and rehabilitation and elaborated a number of principles which we thought held true, not only in the particular hospitals where we worked, but in general. The longer a schizophrenic patient had been in hospital the more he wanted to *stay* in hospital (6). Poverty of the social environment was associated with poverty of clinical condition, and decrease in one kind of poverty was accompanied by decrease in the other (7, 11). Very severely impaired schizophrenic patients responded at once to certain kinds of social stimulation but reverted to their previous level at once when the stimulation was withdrawn (8). Attitudes to work changed when patients were given a course of industrial rehabilitation and this was associated with subsequent work status at follow-up. Attitude to discharge, on the other hand, did not change (5, 9). These principles were drawn from a series of studies, some experimental, some quasi-experimental, and the results should be repeatable. One hopes they are a contribution to knowledge even if they are not immediately translated into a programme of change.

This can be true even of negative results. We looked at three communities which provided different kinds of services for schizophrenic patients outside hospital, but could not find that the course of the illness, the social impairments of the patients, or the problems of the relatives, varied very much in the three areas. We put forward certain suggestions as to why this might have been so (1). Again, there are lessons to be learned even if no conscious use is made of the results in those particular areas.

Thus evaluative work, properly undertaken, is useful irrespective of whether it has an immediate impact, and this is because it is, or should be, dealing with problems of general interest even outside the service context. That is, 'applied' work should also be 'basic' work.

Dr Gardner's second question concerns who is to define the programme objectives. So far as the research programme is concerned, the primary objectives should be defined by the research worker. He will, of course, try to put into as concrete and precise terms as possible what the clinician or administrator is aiming at, in so far as he understands it. Presumably he will not be interested at all unless something about the aims attracts him. The research worker must think and act as an independent scientific investigator: he would be useless if he did not.

In the long run, however, these objectives must come down to one or both of two kinds: the reduction or containment of morbidity in the patient, and the reduction or containment of morbidity in the relatives or others with whom the patient is immediately in contact. There is, of course, also the question of the satisfaction and morale of staff, which is very important, both because it affects the patient and in its own right—but this is not primary in the sense that the other two are.

Moreover, the terms in which these goals are defined must be laid down by the research worker and it should be recognized that indirect evidence is second best. Revans (4) showed in a multi-hospital study that the satisfaction and morale of hospital nurses was correlated with the length of stay of patients—the hospitals with the lowest morale contained patients who stayed longest. He did not show that length of stay was important from the point of view of the patient's illness or satisfaction, though one can guess that it would be. He came closer when he showed that, in a low morale hospital, prescriptions were less efficiently dispensed and administered, but even this is still indirect evidence. Research workers do seem very reluctant to take the next step—that of trying to ascertain whether patients are directly affected.

The third question concerned the funding of schemes with built-in evaluation and feedback of results. Those who hand out money for the development of services and those who fund research programmes should certainly be in touch with one another. Perhaps a proportion of funds for new services should be allocated to research. Certainly the results of research should be pointed out to those who are developing services. I would be rather against any more active intervention than this. Good new ideas in services are usually (though not always) developed by people who are not research workers and it is important not to hinder them too much by harping on someone else's views. On the other hand, it is reasonable to insist, in general, on certain minimum standards if there is the money available to implement the recommendations.

Dr Gardner asks two other questions. One concerns the role of the consumer. Consumer associations have been very active and successful but the idea has not yet been applied very seriously in medicine. The nearest approach is perhaps the parents' societies like our National Society for Mentally Handicapped Children or National Society for Autistic Children. They have sometimes been very successful even though at the risk of some unpopularity with professional bodies, in getting across an unorthodox point of view, and setting up services which public bodies do not provide. Voluntary bodies and consumer associations will and should become very important in medicine, and the research worker should watch for the natural experiments they may present.

Finally, Dr Gardner refers rather gloomily to 'investigators, including those of us present at this conference' who are fearful of planned change in mental health services. 'What we may value most is our autonomy and not the improvement of our ultimate product, mental health services.' This may be too sweeping. The research on workshops originally undertaken by Carstairs, O'Connor, and Rawnsley (2) certainly had a big impact on services, some of it planned. The workers concerned were very pleased about it. Kushlick's work on services for the mentally subnormal in Wessex (3) is a model of how to define needs, plan a service to provide for them and evaluate the success of the new provision. Our own efforts in Camberwell have been rather similar. A psychiatric case-register was set up in anticipation of improvements in services which are gradually being introduced. Evaluation is taking place at many different levels: the case register is the most general and a current experimental study of a rehabilitation workshop is the most specific (12). We are fairly sure that this work will have an impact on planning.

REFERENCES

1. Brown, G. W., Bone, M., Dalison, B., and Wing, J. K. (1966). 'Schizophrenia and social care: a comparative, follow-up study of 339 schizophrenic patients', *Maudsley Monograph No. 17* (London: Oxford University Press).

2. Carstairs, G. M., O'Connor, N., and Rawnsley, K. (1956). 'The organisation of a hospital workshop for chronic psychotic patients', *Brit. J. prev. soc. Med.* **10**, 136–40.

3. Kushlick, A., and Cox, G. (1970). 'Planning services for the subnormal in Wessex', in *Psychiatric Case Registers*, eds. Wing, J. K., and Bransby, R. Department of Health Statistical Report Series No. 8 (London: HMSO).

4. Revans, R. W. (1964). *Standards for Morale: Cause and Effect in Hospitals* (London: Oxford University Press).

5. Wing, J. K. (1960). 'A pilot experiment on the rehabilitation of long-hospitalised male schizophrenic patients', *Brit. J. prev. soc. Med.* **14**, 173–80.

6. —— (1962). 'Institutionalism in mental hospitals', *Brit. J. soc. clin. Psychol.* **1**, 38–51.

7. —— and Brown, G. W. (1961). 'Social treatment of chronic schizophrenia: a comparative survey of three mental hospitals', *J. ment. Sci.* **107**, 847–61.

8. —— and Freudenberg, R. K. (1961). 'The response of severely ill chronic schizophrenic patients to social stimulation', *Amer. J. Psychiat.* **118**, 311–22.

9. —— Bennett, D. H., and Denham, J. (1964). 'The industrial rehabilitation of longstay schizophrenic patients', *Report No. 42* (London: HMSO).

10. —— and Bransby, R. (eds.) (1970). *Psychiatric Case Registers*, Department of Health Statistical Report Series No. 8 (London: HMSO).

11. —— and Brown, G. W. (1970). *Institutionalism and Schizophrenia* (London: Cambridge University Press).

12. —— Wing, L., and Hailey, A. (1970). 'The use of case registers for evaluating and planning psychiatric services', in Wing, J. K., and Bransby, R., op. cit.

2

PETER SAINSBURY

MRC Clinical Psychiatry Research Unit
Chichester, Sussex

I want to refer to some general problems on the evaluation of services and to some particular ones on the points raised by Professor Gardner about our own study.

One of the important but perhaps least-exploited applications of epidemiological methods is the assessment of the operation of services. How does the introduction of extramural facilities affect referral rates, for example, or hospital admission rates? To what extent do they meet needs? Do they, as Dr Gruenberg asks, reduce the prevalence of chronic mental illness in the community, and so on? The neglect of such questions is the more surprising when one considers the scale of the new mental health programmes in the U.S.A. or the new services promoted in this country first by the NHS and then by the Mental Health Act. As the popular emphasis is now on comprehensive schemes of community care, I hasten to forestall misunderstanding by including here the definition of community care we adopted in our work: community care is the extension of the psychiatric services of a district to treat patients outside the mental hospital and the development of an appropriate organization for doing this. In view of Professor Gardner's misgivings I should also add that the type of treatment which is then provided to patients in the community is a separate matter: the fact that in Chichester the therapeutic emphasis was directed towards providing clinical rather than social support to patients and their families—a kind of GP psychiatry—was a *given* feature of the community service that we were looking at.

But as I was saying, the current enthusiasm for community psychiatry has been accompanied by very few rigorously planned attempts to evaluate their effects and their effectiveness. Whereas it is now generally accepted that new pharmacological treatments should be evaluated in accordance with recognized scientific and ethical principles, new services, the clinical effects of which may be just as far-reaching and which may also have substantial effects on others in the community, have not resulted in the formulation of a corresponding set of rules for their assessment. Dr de Alarcon and I tried to begin to remedy this situation by planning our inquiry in such a way as to enable us, firstly, to compare a service that met our criterion of a community one—it treated five-sixths of referrals extramurally—with a control service that admitted one-half of those referred; secondly, to obtain sufficient clinical and social data on each patient to enable the referrals to be matched on relevant variables; and thirdly, to examine *all* referrals to *every* facility from the two defined populations to enable us to make observations of an epidemiological kind on the service as a whole. We believed that the research worker who wishes to evaluate a service will have to combine and avail himself of methods developed in clinical trials and those of the epidemiologist if outcome of the patients' illness as well as the operation of the service are his goals.

Professor Gardner regrets our 'inability to utilize an experimental design'. But what is understood by experimental design in the context of evaluative research? It is necessary to differentiate the circumstances that demand procedures appropriate to experimental research on the one hand and operational research on the other. The investigator using an experimental design deliberately manipulates his experimental population, for example, by random allocation to two differing forms of care. In an operational study, on the other hand, he has to devise methods to investigate what is there; and this means he must avoid any deliberate attempt to alter his population. This, of course, will not prevent him from following the usual precepts whereby he ensures the reliability and validity of his observations; but in this more complex situation he will probably need much more data on all the cases in order to be able to parcel out the effects of the many variables that may be accounting for differences he observes between the two forms of care or services. We therefore did our best to ensure the reliability of our observations and to include data on the pertinent variables as far as we could anticipate them.

Another source of confusion repeatedly encountered when a comparison between services is discussed is the expectation that an evaluation will demonstrate one service as superior to or 'better' than the other, when clearly the real problem and purpose of the evaluator is to determine those social and clinical characteristics of patients and the circumstances of their families that make it possible to predict the type of facility in which a favourable outcome is probable: for example, which patients will do well in hospital and which benefit from home care. It is most unlikely that one kind of care will be consistently superior to another and it is pointless to attempt to approach the problem in this way.

In the time remaining to me I want to consider some problems of method we encountered in evaluating the community psychiatry service in Chichester.

To begin with there is the real difficulty of defining precisely the purpose and policy of the service being assessed—these often tend not to be made explicit by its innovators and so it may fall to the research worker to interpret their aspirations. But once formulated the next step is to select research aims that are few, precise, relevant, and practicable.

Briefly, in Chichester we decided to ask how the introduction of a community psychiatric service had affected, firstly, who is referred to it; secondly, who is treated in hospital and who extramurally; thirdly, what are the effects on the family of treating patients outside of hospital; and lastly, what is the patient's clinical and social outcome.

The next problem is what the criteria for selecting patients should be. Since we aimed to study the effects of the service as a whole, it was necessary to include *all* referrals to it from the catchment area population, as this would allow us to answer the first two questions by obtaining the rates of referral, of admission, and of contact with the different facilities for specific categories of patients. These rates depicted treated prevalence, as there were no psychiatric facilities available in the area other than the service ones.

Another difficulty is selecting and categorizing the items that will by systematically recorded on each patient, and then ensuring that as far as possible they will be objectively defined and reliably obtained. These requirements will tend to limit what can be usefully recorded and will entail abandoning many of one's favourite items.

If the cases appearing within an area are being referred to a number of facilities, the setting up of a *cumulative case-register* will enable basic identifying, personal, clinical, and social data to be obtained in a standard form. This invaluable epidemiological tool will also avoid duplication of cases and allow details of subsequent events and contacts with facilities to be accumulated, which will often be a major consideration in assessing the service. A case-register will also enable random samples to be drawn for more intensive study.

A further problem is that of having some standard against which to compare the service under scrutiny—in other words some form of control. We debated whether or not to compare the data on patients before and after the introduction of the service. For information that is consistently and reliably recorded, such as suicides or dates of admission and discharge, this is an appropriate procedure and one which we used to assess the effect of the service on suicide and on numbers admitted. But for most purposes the earlier data will not have been recorded in the required form, in which case comparisons must be made with another service (or with another part of the same service) that operates an alternative policy. In the event, we chose the Salisbury psychiatric services because the district was very similar demographically, as were the service facilities; but it differed markedly in its policy in that it treated the majority of patients by admission without first screening them.

Since random allocation of patients to two differing services is not a practical proposition, nor can it be done without interfering with what is being studied, comparison is only possible if all psychiatric patients who present for treatment within the two districts are included in the survey, and if enough information is obtained on each to match them on interfering variables and undertake multivariate analyses.

This is particularly the case where assessment of outcome is the goal, when the severity of the patients' illness in the two services must be allowed for. We first compared the referrals to Chichester and Salisbury on severity of illness in the usual way by seeing how well they matched on clinical characteristics. However, we went one step further. Dr de Alarcon obtained a family problem score for each case on a scale whose reliability had been satisfactorily established. Our assumption was that the number of problems, such as the hours spent with patient, his effects on family income, etc., would relate closely to the severity of illness and afford (and in this I agree with Professor Gardner) a more appropriate measure of it than would a diagnostic label or a list of symptoms. I am therefore anxious to assure him that at the time of referral, that is, before intervention by the service psychiatrist, the Chichester and Salisbury patients were surprisingly closely matched on family problem scores, and so, we inferred, on severity of illness.

A last word about the difficult question of measuring outcome. We used many measures of social and clinical outcome: they included ratings by two psychiatrists, by the general practitioner, the patient, the family, the family burden rating, the work record, symptom remission, etc.; but this amount of intelligence has led to problems of its own. We find a better outcome in the community service according to some measures, but not to others; and 'outcome' does not mean the same thing to all observers: psychiatrists and patients, for example, agree with one another, and so do relatives and general practitioners, but the latter concur less with the psychiatrists than between themselves. It appears that there may be different 'dimensions' of outcome which we are now trying to identify with the aid of various factor analytic procedures.

Another aspect of the problem of assessing outcome is to find the best way to deal with the different variables that may be affecting it when random allocation is impracticable. For instance, we recorded in some detail the nature and number of contacts each case made with the social and welfare workers, and consequently were able to show, regardless of service, the extent to which outcome depended on the support that is given to the patient and his family; and concluded, not surprisingly, that success in a community service is clearly associated with the social support that it provides: those patients in whom the families' problems were recognized and acted upon did as well in Chichester as in Salisbury, although a larger proportion were treated at home in the community service; on the other hand, failure to deal with the family problems was accompanied by a less satisfactory outcome in this service. But it was much more difficult to find a satisfactory method of recording the treatment given in each service during two years and so determine how this affected differences in outcome. A rather crude categorization of the forms of treatment prescribed to each patient, such as number of ECT, whether or not he attended occupational therapy, etc., revealed no marked differences between services; nevertheless, this gave us no indication of the quality of care received and *its* effect.

Professor Gardner raises the question of the usefulness of evaluative research. I find it hard to say whether our work had much impact on the development of our service. It has certainly changed; the earlier shortcomings as regards social support to families and patients have been remedied, and the setting up of a new service in the adjoining Horsham district has avoided many of the defects of the Chichester service. But this was not achieved by any process of formal consultation with us, or by including our criticisms and comments on any committee's agenda and then resolving to take action. Nevertheless, I do think that in the course of many discussions with the service doctors and other participants, our findings were in some way assimilated and affected developments.

7

THE USE OF CASE-REGISTERS OF PSYCHIATRIC CARE IN EPIDEMIOLOGICAL RESEARCH OF MENTAL DISORDERS

The use of case-registers of psychiatric care in epidemiological research of mental disorders

ALEX RICHMAN
Psychiatric Epidemiology Research Unit
(New York State Department of Mental Hygiene and Columbia University)
New York, U.S.A.

> We seek the numerical history of each hundred cases of insanity traced from the day when the disease first showed itself, to its resolution in recovery or death. We want to know not merely the age on admission, the duration of the disorder, but farther, its history and progress, with the influence of age, sex, occupation, social relations on the incubation, progress and result of the disease.
>
> ROBERTSON, 1861

This paper is concerned with the use of case-registers in epidemiologic research in mental disorders.[1]

Epidemiologic research on physical or mental illnesses may be based on direct clinical examinations, responses to health questionnaires or data derived from records of contact with health services. Records of contact with health services may be collated to produce case-registers.

1. *Background*

Case-registers were developed by health departments in order to maintain a current summary of medical and health records on persons who required some type of supervision over relatively long periods (45). Case-registers have been developed for conditions such as tuberculosis, cancer, rheumatic fever, and, more recently, for blindness, battered children, narcotic addicts, and persons in psychiatric care.

---

[1] The preparation of this paper has been supported by Grant MH-11105 from the National Institute of Mental Health, and the Mental Health Research Fund of the Canadian Mental Health Association.

Case-registers might cover a local hospital, a number of selected hospitals, or be population-based. A population-based register for cancer would aim at including, for every case in a defined location, detailed information on site, type, stage of disease treatment, and subsequent course. Although other sources might more quickly and cheaply produce results similar to some of those produced by case-registers, the case-register allows one to assess the coverage and reliability of such other sources (33).

The WHO Expert Committee on Health Statistics (53) noted that a population-based register could be an important epidemiologic tool if the majority of cases in the region were registered. It was emphasized that such registration involved more than 'notification' in that a permanent record should be established, the cases followed up, basic statistical tabulations provided on frequency and survival, and that the patients on the register should frequently be the subjects of special studies. In considering various physical diseases of epidemiological interest, the Expert Committee felt that although standardization of diagnosis and comprehensiveness of reporting were major problems, even more serious was the difficulty in defining and accumulating the ancillary data that may be relevant to the disease in question.

Registers for cancer and mental disorder were discussed by the study group in some detail. Cancer was considered highly suitable for register work in that the patient would usually seek medical care and the diagnosis was relatively specific and definitive. Mental disorders illustrated difficulties in diagnosis and in obtaining comprehensiveness of registration; a great variety of non-medical agencies dealt with problems such as alcoholism, and behavioural disturbances in children. The need to develop standard procedures and definitions in case-registers for mental disorders was stressed by the group.

Since the epidemiological usefulness of a register increased the longer it had been in existence, the WHO group considered it necessary to continue careful documentation and quality-control procedures for many years before the register became fully productive for epidemiological purposes. The Committee recommended that caution be exercised in opening new registers for epidemiological purposes and that competently managed registers in a few well-selected areas were likely to be more productive than similar resources devoted to broader but more superficial coverage. The '. . . imbalance between the amount of effort that has gone into the collection of register data and that which has been devoted to its analysis' was also recognized.

2. *Research in the utilization of psychiatric care*

Research in mental disorders has been hampered by the lack of reliable methods for ascertaining disorders in general populations by means of clinical interviews, the absence of valid methods of assessing mental disorder from questionnaire or test responses, and the gap between utilization of psychiatric services and the distribution of morbidity in the population.

The utilization of psychiatric care is the end point of a complex set of social circumstances, involving such factors as geography, finances, the availability of alternatives, attitudes, etc., implanted upon a process of '. . . self, family or community diagnosis based on folk notions and filtered medical ones about what constitutes psychiatric illness and what should be done about it . . .' (7). Studies of persons attending psychiatric facilities must be based on the recognition that only a portion of those with psychiatric illness are likely to attend a physician, to be recognized and recorded as having a psychiatric illness, or to be referred to a psychiatrist for further consultation or care. However, utilization studies deal with the kinds of disorders for which clinical services are provided, are useful in considering the operation of clinical services and provide the means for developing other kinds of epidemiologic approach.

Statistical studies of patients in mental hospitals were initiated early in the history of psychiatry (12, 13, 46). Actuarial methods for obtaining discharge, death, or retention rates were reported by Farr (14). Then, as now, studies of patient movement were hindered by difficulties in defining the severity of disorder and degree of recovery, by lack of uniform nomenclature, and by defects in follow-up. Thurman (46) pointed out that '. . . the average duration of residence may be affected very materially by circumstances altogether independent of the general economy of different institutions, and of the curative methods which are pursued in them'.

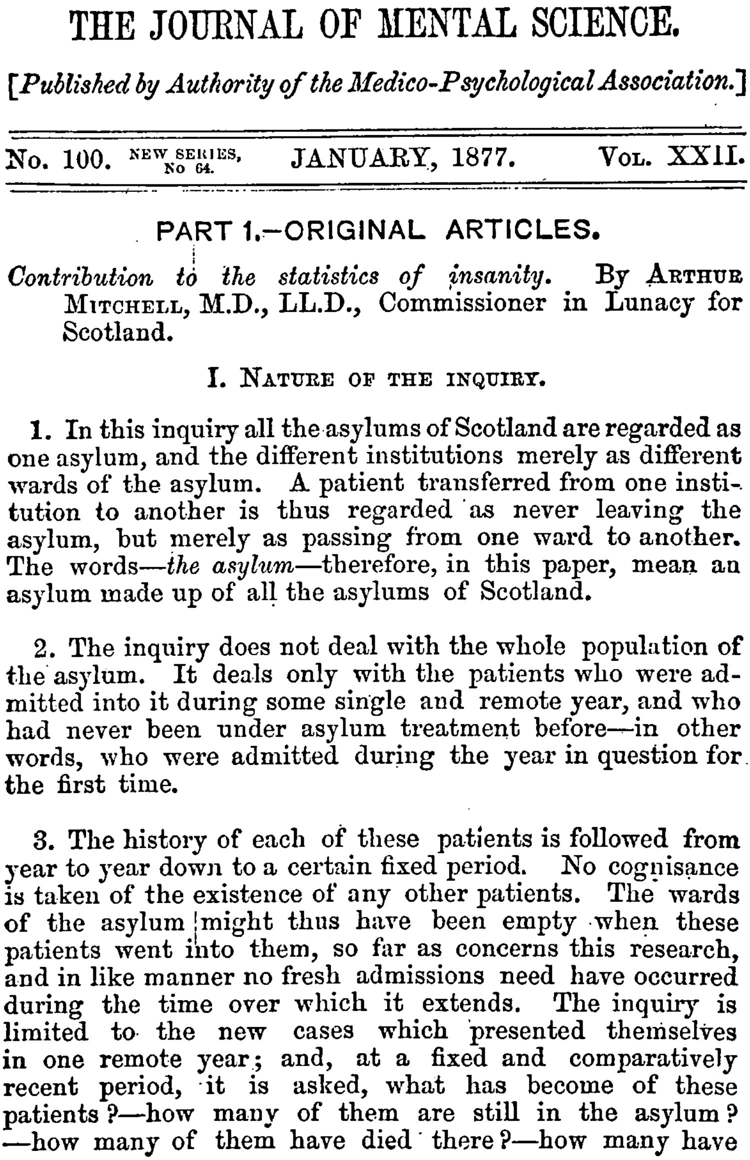
National and international efforts were made to standardize the statistics of mental hospitals, by the Medico-Psychological Association in 1865, the Association of Medical Superintendents of American Institutions for the Insane in 1871, and the International Congress of Alienists.

Within the current century, systematic studies of large-scale mental hospital systems were initiated in New York State by Pollack and Malzberg, in Massachusetts by Dayton, in Canada by the Dominion Bureau of Statistics (36), in the U.S. by Kramer, and in England by Brooke. Kramer (21) stated that it was difficult to understand the mental hospital population unless all the facts were known about the circumstances and forces in the community that bring people into the hospital, and the forces that get patients out of the hospital and keep them out. More community studies of mental illness were needed so that the hospital picture could be put into a meaningful perspective. Studies of mental hospital usage alone presented isolated facts that were quite puzzling and difficult to interpret; it was necessary to study availability and usage of other community resources for diagnosis and treatment of mental disorder (for example, general hospitals with psychiatric treatment services, psychiatric clinics, and private psychiatrists).

3. *Psychiatric case-registers*

Psychiatric case-registers are systems whereby records from a specified set of psychiatric facilities are collected for individual persons from a defined population and accumulated over time. It is of interest to note that such a system was developed for Scottish mental hospitals nearly a hundred years ago (27) (Fig. 7.1).

Although there was considerable interest in the development and analysis of cumulative records of hospital care for individuals, few systems were developed for continuing analysis of such data for defined areas. An adequate registration system was introduced in Norway in 1936, making it possible to follow each individual patient from his first admission to any Norwegian psychiatric hospital through all his successive discharges and readmissions (31). The Norwegian statistical system was kept as simple as possible, while demands for deeper clinical know-



THE JOURNAL OF MENTAL SCIENCE.

[*Published by Authority of the Medico-Psychological Association.*]

No. 100. NEW SERIES, No 64. JANUARY, 1877. Vol. XXII.

PART 1.—ORIGINAL ARTICLES.

Contribution to the statistics of insanity. By Arthur Mitchell, M.D., LL.D., Commissioner in Lunacy for Scotland.

I. Nature of the inquiry.

1. In this inquiry all the asylums of Scotland are regarded as one asylum, and the different institutions merely as different wards of the asylum. A patient transferred from one institution to another is thus regarded as never leaving the asylum, but merely as passing from one ward to another. The words—*the asylum*—therefore, in this paper, mean an asylum made up of all the asylums of Scotland.

2. The inquiry does not deal with the whole population of the asylum. It deals only with the patients who were admitted into it during some single and remote year, and who had never been under asylum treatment before—in other words, who were admitted during the year in question for the first time.

3. The history of each of these patients is followed from year to year down to a certain fixed period. No cognisance is taken of the existence of any other patients. The wards of the asylum might thus have been empty when these patients went into them, so far as concerns this research, and in like manner no fresh admissions need have occurred during the time over which it extends. The inquiry is limited to the new cases which presented themselves in one remote year; and, at a fixed and comparatively recent period, it is asked, what has become of these patients?—how many of them are still in the asylum?—how many of them have died there?—how many have

FIG. 7.1. Report of an early psychiatric case-register in Scotland

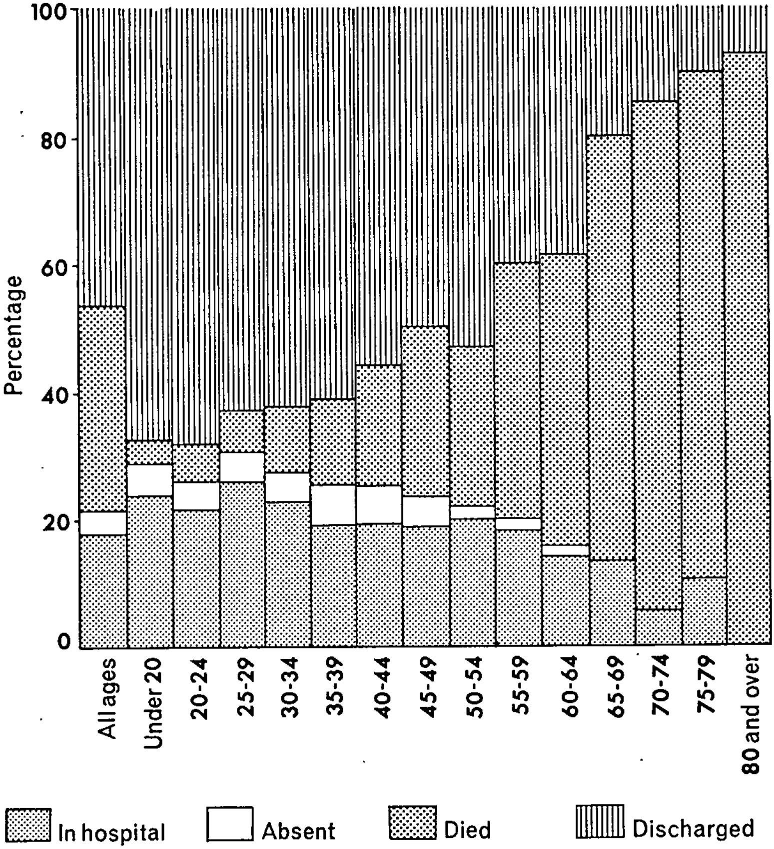


FIG. 7.2. 1952 status of 1944 first admissions, by age, into Ohio Mental Hospitals

ledge were met by intensive studies of representative samples (32). In the U.S., the Ohio Department of Public Welfare began publishing in 1947 the hospital status, at anniversary dates, of first admissions to mental hospitals during successive years. Status included whether the patient had been readmitted to other state hospitals after discharge (41) (Fig. 7.2).

Current emphasis on psychiatric case-registers has thus evolved from long-standing interest in describing the characteristics of individuals using psychiatric services and delineating the way in which facilities were used. Recent advances in data processing and changes in the organization of mental health services have facilitated the linkage and up-dating of records for individuals.

In the past decade population-based case-registers for a variety of in-patient and out-patient facilities have been developed. In the U.S. the National Institute of Mental Health's program of biometric research, sparked by Kramer, led to the development of case-registers in a number of areas. Bahn (4) describes two statewide and five community registers, only one of which was then more than two years old. Outside the U.S.A., case-registers were developed for areas in England (51, 52), Scotland (6), and Denmark (19). A conference on Psychiatric Case Registers was held at the Institute of Psychiatry, London, in 1968 (50).

4. *Some considerations in the operation of psychiatric case-registers*

In the first part of this paper the major problems of case-registers were described as standardization of diagnosis; comprehensiveness of reporting; and the difficulties of defining and accumulating the ancillary data that may be relevant to the disorder. This section describes some of these problems in relation to the operation of psychiatric case-registers.

DIAGNOSIS

Variations in the nomenclature and classification of mental disorders have plagued psychiatric research for many years. Major progress is being made through current studies of variation in diagnostic usage (10), attempts to develop more reliable classifications of mental disorders (42), and the elaboration of standardized interviews for determining the mental status of patients (44, 48).

Changes in diagnosis may occur during the course of care. It is difficult for registers to up-date diagnoses systematically. Frequently patients under continuing care are classified by the diagnosis at admission.

COMPREHENSIVENESS

Comprehensiveness is a crucial problem for psychiatric case registers. Persons with mental disorders are seen in a wide range of community facilities (health, education, welfare, and forensic) which are not primarily psychiatric services (43). There is no reliable method of description, nomenclature, and classification for adequately differentiating mental disorders from other types of psycho-social disorders; nor do the community agencies encompass all of the persons described by Bentley and Cowdry (1934) as '. . . under the care or the advice of . . . healers, magicians, cult leaders, quacks, astrologers, radio advisers, and . . . a great number of the queer, the vagrant, the flighty, the incorrigible, the suspicious, the irascible, the unstable, and the reclusive'. The ultimate extension of the desire for comprehensiveness is represented by Pollack's (34) suggestion that a case-register for the mentally retarded be extended to annual registration of each individual in the population from the time of his birth until death. Such annual registration would include statements of health, education, occupation, achievement, behaviour, and a social record of memberships and affiliations.

There is increasing evidence of the widespread burden of clinically recognizable mental disorder. Hagnell (17), for example, estimated the cumulated risk as being over 50 per cent for developing a mental disorder. However, no matter how many services are included the discrepancy between morbidity and utilization persists.

5. *Ancillary data relevant to psychiatric care*

Contact with a psychiatric facility requires additional data on the reason for the patient's attendance, on the relation between the patient's need for care and the care used, and on the kind and amount of care used.

REASON FOR THE PATIENT'S ATTENDANCE

The reasons for contact with a psychiatric facility have changed as the nature and availability of psychiatric care have been extended. Originally hospitalization was indicated for protection of the patient from himself and of society from the patient. Currently, hospitalization may occur for a variety of reasons (16). Similarly, there may be a wide variety of reasons for attending a psychiatrist on an out-patient basis. Gruenberg (15) described such reasons as including screening examinations for employment, admission to a university or the armed forces; seeing psychiatrists as a source of information about someone who is a patient or is a potential patient; and the appraisal and treatment by psychiatrists of children through contact only with informants rather than the child.

THE NEED FOR PSYCHIATRIC CARE

Most persons who consult a psychiatrist are considered to need further care. Avnet (3) described an experiment wherein subscribers to a prepaid medical insurance plan could obtain a psychiatric consultation for $5 upon self-referral; 'practically every patient who crossed the psychiatric threshold was judged at the initial interview to be in need of treatment and was treated if he returned'.

The care to which the patient is admitted may not be that needed by the patient. That is, the patient may have been hospitalized because of a lack of the required form of care in the community. Mendel and Rapport (26) found that in their setting one-third of the emergency patients hospitalized between 5 p.m. and 8 a.m. could be discharged the next morning and concluded that this was an indication of the unavailability of alternative support resources in the community during those hours. The Maryland Case Register provides for reporting discharges in terms of 'other community resource needed but not available'. It is more difficult to get adequate description of the patient's needs for care at the time of admission.

KIND OF CARE

It is increasingly difficult to assess the kind of care represented by admission to a particular facility. The discharge of a patient the morning after admission represents a different type of care from a longer hospital stay. It is thus not possible to infer the types or amount of care associated with admission.

6. *Descriptive statistics on the utilization of patient care*

When Gertrude Stein was on her death-bed, one of her friends asked her 'What is the answer?' After a few seconds she whis-

---

A. EVENTS (admissions, discharges, deaths).

B. PERSONS (individuals receiving care during a specified interval).

C. EVENTS REPORTED FOR INDIVIDUAL PERSONS
  i. Two consecutive events, e.g. discharge of admissions, readmissions of discharges.
  ii. Two events defined by time, e.g. status one year after admission.
  iii. All events occurring during a specified interval, e.g. number of in-patient days, persons receiving out-patient care only.

D. TREATMENT STATUS OF POPULATION AT A GIVEN POINT OF TIME
  i. Under care
  ii. Not under care: previous care reported
                no previous care reported.

FIG. 7.3. Case-register data on the utilization of psychiatric care

pered back 'What is the question?' (*Encyclopaedia of Social Science*).

Case-registers provide data about the cumulated sequence of admissions and discharges reported by psychiatric facilities for a group of unduplicated persons. Such data may be used to supply answers to various kinds of questions (Fig. 7.3).

Epidemiologic distributions of persons entering or receiving care may be determined but additional data are required in order to interpret such admission or prevalence rates in terms of the kinds, duration, amount, or patterns of care they represent. The prevalence of patients with specified conditions is not necessarily a true reflection of the needs for facilities or personnel (Commission on Chronic Illness). On the one hand the number of persons represents the size of the burden on the community of those for whom some unspecified form of psychiatric care may be needed, or, on the other hand, they indicate the proportion of the communities resources devoted to psychiatry (52).

Data on events for individuals provide perspectives on the duration of care used and make it possible to make inferences on the severity of the disability from the duration of care. Data on two consecutive events give some perspective on the duration of care, mortality, or readmission rates. Such data describe one link of the chain. Without data on the severity of the disorder at admission and the degree of subsequent change, it is difficult to distinguish to what extent shorter durations of stay are associated with mild disorders or effective treatment.

The lack of reported out-patient care following discharge from mental hospitals may represent a failure of follow-up, a high degree of recovery, or the receipt of services from non-psychiatric agencies.

7. *Alternative sources of data*

Case-registers represent the only way for accumulating data on all types of psychiatric care for individuals. However, data on certain kinds of psychiatric care can be obtained for individuals from other sources.

Transaction files of state hospital systems can provide data on individuals in hospital at a point in time. In addition they may be used to describe the attrition of long-stay residents, the number and characteristics of persons entering the long-stay group (38), the hospital treatment status at the end of the year for patients in hospital at the start of the year, and the actual number of days spent in hospital by a group of patients continuously 'on the books' but intermittently in hospital, on convalescent care, or on leave.

Some prepaid medical insurance plans provide comprehensive care including psychiatric consultation and psychotherapy. Several of the utilization data collected for administrative purposes by the plan are suitable for assessing the duration and the amount of psychiatric care (36).

A census of patients in Danish psychiatric institutions was initiated in 1957, and has been repeated at five-year intervals. This type of data allows one to compare hospital status of individuals at the beginning and end of the five-year interval (19).

8. *Psychiatric case-registers in the evaluation of community mental health services*

Playwright with strong plot, strong characters, strong message and weak writing skills looking for collaborator. Collaborator should be playwright with outstanding skills looking for strong plot, strong characters, and strong message.

Advertisement, *Village Voice* (New York) 4 April 1968

Psychiatric case-registers make their major contribution to the evaluation of community mental health services when there is a clear recognition of the use of the register in evaluation and planning. Counts of unduplicated persons with mental disorders under psychiatric care are not the only need of persons planning or evaluating community mental health programs (15). Wing describes evaluation as ideally requiring answers to the following six topics.

(1) How many individuals are in contact with existing services?

(2) What are their needs and those of their relatives? That is, what are their symptoms and disabilities and what burdens are placed upon relatives and the rest of the community?

(3) Are the services at present provided meeting these needs?

(4) How many others, not in touch with the existing services, also have needs?

(5) What new services, or modifications to existing services, are required to cater for unmet needs?

(6) When these innovations are introduced, do they in fact meet these needs?

Psychiatric case-registers can best provide data on the first question. Additional or other data are required to answer the remaining questions, for 'no investigation of administrative indices such as length of stay or readmission rate can replace the fundamental need directly to measure morbidity in the patient and his family' (48).

It is hard to foretell the future of mental health programs, since we do not know the relation between their present utilization and morbidity in the population. Care is not now uniformly available for all types of psychiatric illness. Conditions for which patients were more likely to receive care will not increase to the same extent as those conditions for which patients were less likely to receive care. However, one can obtain an indication from the case-register of disorders not being reported, of conditions representing failures in the use of existing knowledge regarding prevention and treatment, and the distribution of different types of care.

Case-registers are invaluable when linked with long-term research programs in the epidemiology of mental disorders. To paraphrase Morris, case-registers are the only way of asking some questions, one way of asking others, and no way at all to ask many.

REFERENCES

1. ACHESON, E. D., and BARR, A. (1965). 'Multiple spells of in-patient treatment in a calendar year', *Brit. J. prev. soc. Med.* **19**, 182–91.
2. ADELSTEIN, A. M., DOWNHAM, D. Y., STEIN, Z., and SUSSER, M. W. (1968). 'The epidemiology of mental illness in an English city', *Social Psychiatry*, **3**, 47–59.
3. AVNET, H. H. (1962). *Psychiatric Insurance* (New York Group Health Insurance).
4. BAHN, A. K. (1962). 'Psychiatric Case Register Conference, 1962', *Public Health Reports*, **77**, 1071–6.
5. —— (1964). 'Experience and philosophy with regard to case registers in health and welfare', presented at Workshop on Case Registers, American Orthopsychiatric Association, Chicago, 1964.
6. BALDWIN, J. A., INNES, G., MILLAR, W. M., SHARP, G. A., and DOSSICOTT, N. (1965). 'A psychiatric case register in north-east Scotland', *Brit. J. prev. soc. Med.* **19**, 38–42.
7. BLUM, R. H. (1962). 'Case identification in psychiatric epidemiology: methods and problems', *Milbank mem. Fd Quart.* **40**, 253–88.
8. BROOKE, E. M. (1963). 'A cohort study of patients, first admitted to mental hospitals in 1954 and 1955', *General Register Office Studies on Medical Population Subjects, No. 18* (London: HMSO).
9. COMMISSION ON CHRONIC ILLNESS. (1957). *Chronic Illness in the United States*, vol. iv, *Chronic Illness in a Large City* (Cambridge: Harvard University Press).
10. COOPER, J. E., KENDALL, R. E., GURLAND, B. J., SARTORIUS, N., and FARKAS, T. (1969). 'Cross-national study of diagnosis of the mental disorders: some results from the first comparative investigation', *Amer. J. Psychiat.* **125**, 21–9, April supplement.
11. DAYTON, N. A. (1940). *New Facts on Mental Disorders* (Springfield, Ill.: Charles C. Thomas).
12. EARLE, P. (1848). *History, Description and Statistics of the Bloomingdale Asylum for the Insane* (New York: Egbert, Hovey and King).

13. Esquirol, J. E. D. (1838). *Des Maladies mentales considérées sous les rapports médical, hygiénique et médico-légal* (Paris: J. B. Bailliere).

14. Farr, W. (1841). 'Report upon the mortality of lunatics', *Royal Statistical Society Journal*, **4**, 17–33.

15. Gruenberg, E. M. (1968). 'Epidemiology and medical care statistics', *The Role and Methodology of Classification in Psychiatry and Psychopathology* (Washington: USGPO).

16. Group for the Advancement of Psychiatry (1969). *Crisis in Psychiatric Hospitalization*, Vol. 7, No. 2 (New York: The Association).

17. Hagnell, O. (1966). *A Prospective Study of the Incidence of Mental Disorder* (Norstedts: Svenska Bokforlaget).

18. Juel-Nielsen, N., Bille, M., Flygensing, J., and Helgason, T. (1961). 'Frequency of depressive states within geographically delimited population groups', *Acta psychiat. Scand.*, Suppl. **162**, 37–69–80.

19. —— and Stromgren, E. (1963). 'Five years later', *Acta Jutlandica*, **35**, 1.

20. Kramer, M. (1953). 'Long range studies of mental hospital patients: an important area for research in chronic disease', *Milbank mem. Fd Quart.* **31**, 253–64.

21. —— (1954). 'Proceedings of the Third Conference of Mental Hospital Administrators and Statisticians', *Public Health Service Publication No. 348* (Washington: USGPO).

22. Le Riche, H. (1960). 'A linkage of data on physicians' services and hospital separations for 1960 in Ontario', *Special Report No. 2* (Toronto: Physicians' Services Inc.).

23. Lin, T.-Y. (1967). 'The epidemiological study of mental disorders by the WHO', *Social Psychiatry*, **1**, 204–6.

24. Malzberg, B. (1956). 'Cohort studies of mental disease in New York State, 1943–9', *Ment. Hyg. (N.Y.)*, **40**, 450–79.

25. Medico-Psychological Association (1865). 'Proceedings of the annual meeting, 13 July 1865. Report of the Committee upon Asylum Statistics', *J. ment. Sci.* **11**, 402–8.

26. Mendel, M. M., and Rapport, S. (1969). 'Determinants of the decision for psychiatric hospitalization', *Arch. gen. Psychiat.* **20**, 321–8.

27. Mitchell, A. (1877). 'Contribution to the statistics of insanity', *J. ment. Sci.* **22**, 507–15.

28. Morris, J. N. (1964). *Uses of Epidemiology* (Edinburgh and London: E. S. Livingstone Ltd.).

29. National Institute of Mental Health. (1968). *Manual of Standard Reporting Items for Psychiatric Case Register* (Register Development and Studies Section, May).

30. Newcombe, H. B., Kennedy, J. M., Axford, S. J., and James, A. P. (1959). 'Automatic linkage of vital records', *Science*, **130**, 954–9.

31. Odegard, O. (1968). 'The pattern of discharge and readmission in Norwegian mental hospitals', *Amer. J. Psychiat.* **125**, 333–40.

32. —— (1961). 'Current studies of incidence and prevalence of hospitalized mental patients in Scandanavia', in *Comparative Epidemiology of the Mental Disorders*, eds. Hoch, P., and Zubin, J. (New York: Grune and Stratton).

33. PEDERSEN, E. (1962). 'Some uses of the cancer registry in cancer control', *Brit. J. prev. soc. Med.* **16**, 105–10.

34. POLLOCK, H. M. (1947). 'Annual registration of the entire population', *Amer. J. ment. Defic.* **51**, 731–4.

35. —— (1945). 'Development of the statistics of mental disease in the United States during the past century', *Amer. J. Psychiat.* **102**, 1–17.

36. RICHMAN, A. (1966). *Psychiatric Care in Canada: Extent and Results.* A study prepared for the Royal Commission on Health Services, 1964 (Ottawa: Queen's Printer).

37. —— and TYHURST, J. S. (1957). 'Patient movement in psychiatric hospitals', read at annual meeting, Canadian Psychiatric Association, 21 June, Edmonton, Alta.

38. —— and WIEHL, D. G. (1968). 'Mental hospital stay beyond two years: trends for admissions from Manhattan 1963–8', presented at annual meeting, American Public Health Association, Detroit.

39. ROBERTSON, L. L. (1861). 'Suggestions toward a uniform system of asylum statistics', *J. ment. Sci.* **7**, 195–211.

40. ROBERTSON, N. L., BALDWIN, J. A., and HALL, D. J. (1969). 'Psychiatric record linkage: a pilot study', *Research Report No. 9* (University of Aberdeen).

41. SIMMONS, M. B. (1952). 'Status at yearly intervals after admission of patients first admitted to hospitals for the mentally ill during 1944–50', *Public Welfare Statistics (Ohio)*, 7, **5**, 4–10, May.

42. SHEPHERD, M., BROOKE, E. M., COOPER, J. E., and LINN, T. (1968). 'An experimental approach to psychiatric diagnosis', *Acta psychiat. Scand.*, Suppl. 201.

43. SOUTHARD, E. E., and JARRETT, M. C. (1922). *The Kingdom of Evils* (London: George Allen and Unwin).

44. SPITZER, R. L., FLEISS, J. L., BURDOCK, E. I., and HARDESTY, A. S. (1964). 'The mental status schedule: rationale, reliability and validity', *Compr. Psychiat.* **5**, 384–95.

45. SWAROOP, S. (1960). *Introduction to Health Statistics* (Edinburgh and London: E. S. Livingstone Ltd.).

46. THURNAM, J. (1845). *Observations and Essays on the Statistics of Insanity and on Establishments for the Insane: to which are added the statistics of The Retreat, York* (London: Gilpin and Hunton).

47. WALLIS, W. A., and ROBERTS, H. V. (1956). *Statistics: A New Approach* (Glencoe: The Free Press).

48. WING, J. K. (1968). 'Patients with psychiatric disorder', *Research Seminar on the Evaluation of Mental Health: an International Perspective.* (San Francisco: Jossey-Bass Inc.).

49. —— BIRLEY, J. L. T., COOPER, J. E., GRAHAM, P., and ISAACS, A. D. (1967). 'The reliability of a procedure for measuring and clarifying Present Psychiatric State', *Brit. J. Psychiat.* **113**, 499–515.

50. —— and BRANSBY, R. (1969). *Psychiatric case registers*, Department of Health Statistical Report Series No. 8 (London: HMSO).

51. WING, L., BRAMLEY, C., HAILY, A. and WING, J. K. (1968). 'Camberwell cumulative psychiatric case register. Part I: aims and methods', *Social Psychiatry*, **3**, 116–22.

52. —— Wing, J. K., Haily, A., Bahn, A. K., Smith, H. E., and Baldwin, J. A. (1967). 'The use of psychiatric services in three urban areas: an international case register study', ibid. **2**, 158–67.

53. World Health Organization. (1967). 'Epidemiological methods in the study of chronic diseases', *Technical Report Series no. 365*, *Expert Committee on Health Statistics* (Geneva: WHO).

Discussion

J. A. BALDWIN

Medical Director, Oxford Record Linkage Study and Unit of Clinical Epidemiology, University of Oxford
(formerly of the Department of Psychiatry, Aberdeen University)

In north-east Scotland we have had a case-register fully operational for six and a half years, covering every type of psychiatric contact for our population of nearly half a million. We think of the register as a systematic, cumulative file of linked personal health records which we can use as longitudinal, cross-sectional, person-based, family-based, or event-based data with the utmost flexibility. In terms of its value as a planning and management tool, as a facilitator of clinical research, and as a sampling base for epidemiological and cohort research, the register has amply rewarded the effort put into its development, maintenance, and quality control.

I should like to recall why we needed a case-register, because we had some reasons Dr Richman did not mention. Thus:

(1) Hospitalization statistics were no longer representative of specialist-treated psychiatric morbidity.

(2) We had to have a source of routine, systematic statistics to monitor the function of all the psychiatric services and their interrelations in a dynamic way.

(3) We wished to study *trends* over the years in a rapidly changing situation, both in the use of different types of services and in the pattern of referral rates.

(4) With a small population like ours, the numbers of cases in any single year were not enough to study in sufficient detail many of the questions we wanted to answer, so we needed a number of years of information. We are now working on a five-year incidence file which will permit much more detailed analyses.

(5) We have an interest in long-term follow-up studies, for which at least twenty-year spans are necessary. In this area, with 95 per cent Scottish born and 86 per cent in the region, we have a specially favourable situation for minimizing cohort losses. Incidentally, it is not an argument against registers that they only collect data while patients are in care: there is no need for them to be so limited if community follow-up data are required. Most registers link with death certificates anyway.

(6) The core of the idea of the register is, of course, record linkage, and in this context we were particularly interested in family record linking—married pairs, siblings, and parent–child clusters. The register has proved its value for this purpose since we can collect enough data to make the links and maintain the file over the necessary time periods.

(7) The register has also proved itself an excellent means of encouraging and facilitating clinical as well as specialized epidemiological research by (i) greatly simplifying retrospective searching for specific types of case and prospectively gathering special samples, (ii) testing representativeness of samples, and (iii) obtaining properly matched controls.

The uses we have made of the register are now too many to do justice to in a few minutes. The point about many of these is not that they *could* not otherwise have been done but that they *would* not have been done had the presence of the register not made them practical possibilities. They were possible for us because we could thereby avoid the paraphernalia of grant-getting, staffing, and costly special data collection. Dr Richman has said that registers are an expensive way of doing research. The contrary is true. It has been shown several times, both in Canada and here, that the essence of the register file, the business of linking records, is surprisingly cheap. For the addition of a few items of linkage information, data which are usually collected anyway can be made infinitely more valuable by serving a wide range of additional purposes. In the rest of medicine, record linkage on a national scale is being accepted and implemented because it is cheap and effective. Psychiatry should not drag its feet just because so much psychiatric illness is long-term and because we therefore have to wait a long time for the ultimate value of the register to be realized. Psychiatric data-banks—unlike money banks—appreciate in value in proportion to their longevity.

Lest I appear too enthusiastic, let me end with three points of caution mentioned by Dr Richman, but which bear amplification:

(1) Of course data must be collected for specific purposes. For any worthwhile degree of sophistication in research, much more than the basic common information which can be collected routinely in a register must be obtained and linked with the existing record. Registers which do not have this capability have missed the point. It is true that so far results from registers have not always been as illuminating as some expected in the first flush of enthusiasm; but that is because they have not existed long enough yet, and the methods of analysis, particularly of longitudinal data, are not yet developed enough to exploit fully even the already existing potential.

(2) Registers are not new in medical research and, as Dr Richman pointed out, cohort studies involving record linkage in mental disease were undertaken, at least in Scotland, a century ago. But there have been innumerable cases of failure and even some of misleading results due to inadequate quality control.

(3) The psychiatric case-register is, by definition, restricted to specialist-treated morbidity and hence nine out of ten or nineteen out of twenty cases of psychiatric illness, however defined, are not accounted for. The time is coming when it will be necessary to develop registers of whole populations, perhaps with standard screening procedures incorporated into data collection, to provide baselines against which to measure changes. We have begun to think about this in the Oxford studies, where the new city of Milton Keynes may provide an appropriate experimental area, as it grows from its present population of 40,000 to 250,000 by the end of the century.

2

LORNA WING

MRC Social Psychiatry Research Unit,
Institute of Psychiatry, University of London

Dr Richman brought out some of the problems and limitations of registers as well as their uses: he quoted the warning from the WHO report concerning the tendency to an imbalance between the collection of data on the one hand and its use on the other. This kind of problem arises when a register is set up without there being a pre-existing need for such an expensive instrument. Where a register is developed in an area because it is needed to solve specific problems, the material collected is less likely to be wasted.

This point can be illustrated by the Camberwell Register, which was started in an area in South London where the Medical Research Council's Social Psychiatry Unit had already been working for some years and where there were plans for developing a comprehensive community service for psychiatric and subnormal patients. Camberwell is an area which, prior to the development of an integrated service, was served by a large number of different agencies, all of which took patients from other areas as well. A substantial minority of people attended more than one agency during one episode of illness. In order to investigate and evaluate the effects of changes in the services it was necessary to build up a register based on the area, covering all agencies, which collated information about individual patients and which was cumulative over time. The national statistics could not be used because up to the present they have been based on events, not individuals, and, except for in-patients, only crude data are available.

The evaluation of psychiatric services poses a number of questions (2).

(1) How many and what kinds of individuals are in contact with existing services?

(2) What are their needs, that is, what are their symptoms and their primary and secondary handicaps; and what burdens do these place on the relatives and the rest of the community?

(3) Are the services meeting these needs?

(4) How many other people, not in touch with services, also have needs? Do their needs differ from the needs of people who are seen by psychiatrists?

(5) What new services, or modifications of existing services, are likely to cater for unmet needs?

(6) When innovations are introduced, do they meet these needs?

Dr Richman has already referred to the ways in which registers can help to answer these questions. Some examples of these can be given as they are being applied in Camberwell.

Routine descriptive statistics have been used to look at changes in the distribution of inpatient services in the area (3). The Register started at the end of 1964 with a census of all patients in contact with in-, out-, and day-patient services; and details of all admissions to hospital since the census day have been recorded. The routine tables giving the characteristics of patients admitted to each agency each year have shown differences between these agencies. For example, during the first two years of the register the Maudsley Hospital (a postgraduate teaching hospital) admitted a disproportionately small number of schizophrenic patients and an excess of patients with depression and neurosis as compared with the local area mental hospitals.

In 1967 various planned changes occurred in the local services, including the opening of a new unit for short- and medium-stay in-patients and the adoption of Camberwell as its catchment area by the Maudsley Hospital. The routine tables began to show a fall in admissions to the area mental hospital and a rise in admissions to the new unit and to the Maudsley Hospital. However, it now looks as if the *total* number of short- and medium-stay patients is increasing and that more patients with milder non-psychotic illnesses are being admitted than previously. This is an interesting trend which will need further analysis. The routine statistics are also being used to monitor the changes in the long-stay population of in-patients, and the development of the out-patient services, which are considerably more plentiful in Camberwell than the average for England and Wales.

It is worth underlining the point already made by Dr Richman that a description of a trend gives no information at all as to the effect it has on the patients involved. An evaluation of a service, as distinct from a statistical description, must involve more detailed studies.

The Camberwell register was used as a basis for collecting further information when the specialized psychotherapy and the supportive psychotherapy services in the area were examined. Additional details about patients attending in a specified time period were collected from case-notes and collated with the routine material. From this it was calculated how many extra patients could theoretically have been referred for specialized or supportive psychotherapy and how many extra doctors this extension of the service would need. These calculations, which gave an extremely high estimate of the number of psychiatrists necessary for this additional work, emphasized the point that with money and resources as scarce as they are, decisions about extending or providing new services should be based on a knowledge of their effectiveness (1).

Some other studies have been done involving the collection of additional material for a limited time. These include a comparison of the work of the mental welfare officers, hospital social workers, and joint appointment social workers, and also a study of the severely subnormal children in the area.

The register has been used as a sampling frame for a detailed examination of certain aspects of the local services. One of these studies concerned all the schizophrenic patients contacting the Camberwell services during a specified time period. Without the register, it would have been almost impossible to have obtained a complete coverage of all the patients involved and quite impossible to have collected details of their service experience over three to four years and collated it with the rest of the information gathered in the study.

Another detailed project is the evaluation of a day-hospital and rehabilitation workshop newly opened in the area. The register is used to generate a study population fitting certain criteria of age, diagnosis, time of contact with services, and so on. Further information about work record is taken from the case-notes and then suitable patients are interviewed and tested to determine whether they need a course of rehabilitation. Finally they are placed at random in the experimental or the control group. At the end of the study it should be possible to say (*a*) how many patients of the kind selected are to be found in Camberwell, (*b*) how many complete a course of rehabilitation, (*c*) what is the outcome, (*d*) which patients do best, and (*e*) which aspects of the rehabilitation programme are most useful. This information should make it possible to plan the future rehabilitation needs of Camberwell and the results could be applied, cautiously, to other areas.

The final use of a register which Dr Richman mentioned—that is, as a starting point for the study of people who do not contact psychiatric services—has not been put into effect as yet, but will be important in the future.

There is no doubt that registers can turn into expensive encumbrances, but if used properly they can prove invaluable aids. They should be regarded as pieces of apparatus or instruments analogous to the polygraphs of physiologists and not as ends in themselves.

REFERENCES

1. WING, J. K., and WING, L. (1970). 'Psychotherapy and the National Health Service', *Brit. J. Psychiat.* **116**, 51–6.

2. —— —— and HAILEY, A. (1969). 'The use of case registers for evaluating and planning psychiatric services', in *Psychiatric Case Registers*, eds. Wing, J. K., and Bransby, R. (Department of Health and Social Security Statistical Report Series, London) (to be published).

3. WING, L., WING, J. K., HAILEY, A., and TAYLOR, C. (1969). 'Camberwel Psychiatric Case Register, 31 December 1965–30 June 1968: tables and brief annotations, ibid.

8

PSYCHIATRIC EPIDEMIOLOGY AND GENERAL PRACTICE

Introduction

MICHAEL SHEPHERD
Institute of Psychiatry, University of London

For the past twelve years a small research team at the Institute of Psychiatry in the University of London has been engaged in studying psychiatric illness in the community, with the help of a large number of active and willing general practitioners. In the working conditions of the U.K., the family doctor practising front-line medicine is in a unique position to furnish information about the health and sickness of most members of the population. In this role he is potentially able to collaborate in tackling one of the most difficult problems confronting epidemiological inquiry in the field of mental disorder, namely the detection of a psychiatric 'case' (1). For operational purposes, the 'case' may be defined as the patient whose symptoms, distress, or discomfort lead to a medical consultation at which a psychiatric diagnosis is made by a qualified physician. This notion clearly implies acceptance by the patient, and confirmation by the doctor, of what Professor Mechanic would call 'illness-behaviour'; it is also limited by the boundaries of what we would call 'conspicuous' morbidity. Nonetheless, it has stood the test of an initial series of rigorous inquiries, which we described in some detail three years ago (2).

Two of the findings stand out with particular prominence. The first is the sheer size of the burden of extramural psychiatric illness: of some 15,000 patients at risk during a twelve-month period, just over 2,000, or approximately 14 per cent, consulted their doctor at least once for a condition diagnosed as largely or entirely psychiatric in nature. The second relates to the management of these disorders: it emerged that only about one in twenty of the patients identified in the survey had been referred to any of the mental health facilities despite what the GPs freely acknowledged to be the unsatisfactory nature of the treatment they were able to provide. Clearly there is here a large area of unmet need, with obvious implications for clinical and administrative action.

For such action to be rationally based, however, much more factual information is required, and this has been the objective of the second phase of our programme. The first phase, being concerned in large measure with the acquisition of extensive data on a large population, depended inevitably on the ability of the GP to screen and detect morbidity; the calibration of his capacity to do so constituted an important methodological issue. In the second phase of our work we have been concentrating more on the intensive study of high-risk subgroups and for this purpose measuring instruments have had to be constructed. We have come to rely less on psychiatric assessments made by GPs and have devised methods of case-identification that are independent of the GPs' assessments. The first of these is a screening questionnaire aimed at detecting minor psychiatric illness which has a known high reliability and validity, and which has been shown to be very much more satisfactory than the CMI for purposes of case-identification. The questionnaire is intended to be used in conjunction with a standardized interview administered by a psychiatrist, and it is the design and construction of this instrument which is the substance of Dr Goldberg's presentation. The use of such an instrument is then illustrated by the subsequent two papers. Dr Eastwood has been following up the evidence in our earlier studies which suggested that patients with psychiatric disorder have a higher than average expectation of general morbidity. Dr Cooper summarizes a study inspired by the demonstration that more than half of the psychiatric patients in our original survey could be regarded as suffering from chronic illness in that they had been continuously unwell for at least one year, and so might be expected to exhibit social as well as medical problems.

Between them these three studies will, I hope, serve to demonstrate the importance of examining this large area of sickness by the application of epidemiological methods.

REFERENCES

1. SHEPHERD, M., and COOPER, B. (1964). 'Epidemiology and mental disorder: a review', *J. Neurol. Neurosurg. Psychiat.* **27**, 277.
2. —— —— BROWN, A. C., and KALTON, G. (1966). *Psychiatric Illness in General Practice* (London: Oxford University Press).

The reliability of a standardized psychiatric interview suitable for use in community surveys

D. P. GOLDBERG
Institute of Psychiatry, University of London

The requirements for a standardized psychiatric interview suitable for use in community surveys are rather different from those necessary for the assessment of identified psychiatric patients, which have been described to us by Dr Wing and Dr John Cooper. We have designed an interview that is administered by an experienced psychiatrist in a realistic clinical setting and which gives special emphasis to those types of psychiatric disturbance that are commonly encountered in the community (1). It is relatively economical of time and has been found to be acceptable to patients who do not see themselves as emotionally ill. It generates information about individual symptoms and signs of mental illness as well as allowing the psychiatrist to make an over-all diagnostic assessment, and it discriminates clearly both between mentally disturbed and normal individuals, and between patients with different degrees of psychiatric disturbance. Since it measures different degrees of disturbance between individuals, it is also possible to use it to measure change in a given individual over time (Table 8.1).

The interview with the patient is in three sections and there is then a fourth section of the interview schedule which is completed after the patient has left the room. The first section is an unstructured interview with the patient which is used both to establish rapport and to discover the patient's present symptoms. Any relevant past history is collected at this stage and this part of the interview usually only takes a few minutes. The second part of the interview is a detailed and systematic inquiry into symptoms that the patient may have experienced in the previous week. The third part of the interview is again

TABLE 8.1. *Over-all plan of the clinical interview*

1. Front sheet for name, demographic data, etc.
2. UNSTRUCTURED INTERVIEW
   History of present condition
   Past history (medical and psychiatric)
3. SEMI-STRUCTURED INTERVIEW
   Reported symptoms in the past week, divided into ten sympton areas.
4. UNSTRUCTURED INTERVIEW
   (Only given if indicated)
   Family psychiatric history
   Brief personal and social history
5. Pages for rating abnormalities seen at interview, writing a short formulation and making a diagnosis

unstructured and leaves the interviewer free to collect as much or as little information about the family psychiatric history and personal history as he feels necessary to make his clinical assessment. After the patient has left, the interviewer assesses the manifest abnormalities that he has seen and heard at interview, using twelve rating scales, and then writes a short formulation of the case and makes a clinical diagnosis, using the ICD where this is appropriate.

The second part of the interview, where the patient is asked about the symptoms that have been experienced in the past week, is concerned with the following ten symptoms: somatic symptoms, fatigue, sleep disturbance, irritability, lack of concentration, depression, anxiety, phobias, obsessions and compulsions, depersonalization.

This semi-structured interview begins with symptoms commonly encountered in everyday life, such as headache and fatigue, only gradually moving towards more obviously psychiatric items such as depression and worry. The interviewer does not ask about florid psychotic phenomena unless he suspects that they are present and, in the same way, cognitive and memory testing is only carried out if there is reason to suspect intellectual impairment.

Within each symptom area there are mandatory questions which all patients must be asked and then a series of probes that are only asked if it seems necessary. Table 8.2 shows an example of this part of the interview. In the example the mandatory question is shown above the horizontal line. The probe questions can be varied to suit the individual concerned, the interviewer's task being to establish the frequency, duration, and intensity of the symptom concerned during the previous week. There are simple rules relating the frequency and intensity of each symptom to the 5-point scale on which it is rated. The rating ‘0’ means that the symptom was absent and the rating ‘1’ that although present it did not cause significant distress, or was perhaps a personality trait of the patient. Ratings 2, 3, and 4 are based on the frequency and intensity of the symptom. This 5-point scale was devised with two aims in view: it is relatively clear and simple, and so we have found inter-rater agreement to be high: and it allows us to measure change in a symptom over time.

TABLE 8.2. *A sample page—irritability*

Do you find that you are easily upset or irritable with those around you?

How long have you been like this?
Are you like it all the time, or just occasionally?
What sort of things upset you?
How has it been in the past week?
Have you had any rows with anyone?
Are there still any hard feelings?
Irritability rating: 4 3 2 1 0

Abnormalities seen and heard at interview are rated on twelve scales: three scales for abnormality of behaviour, four scales for abnormal moods, and five scales for perceptual and cognitive abnormalities (Table 8.3). Each scale uses a 5-point rating which is broadly similar to that described above for the reported symptoms. An important difference between the reported symptom ratings and these manifest abnormality ratings is that in the former the interviewer accepts what the patient

TABLE 8.3. *Evidence for psychiatric disturbance noted at interview 'The 12 Manifest Abnormalities'*

1. Slow, lacking spontaneity
2. Suspicious, defensive
3. Histrionic
4. Anxious, agitated, tense
5. Elated, euphoric
6. Flattened, incongruous
7. Depressed
8. Depressive thought content
9. Excessive concern with bodily functions
10. Delusions, thought disorder, misinterpretations
11. Hallucinations
12. Intellectual impairment

says at its face value, while in the latter he is making ratings of his own clinical assessments of the degree of clinical disturbance present. It follows that from the point of view of case-identification it is these ratings that are of especial importance, and it is possible to use weighting procedures to assist in computing an over-all severity score which correlates best with clinical impressions.

Thus the information available about each patient contains ten ratings of symptoms in part 2, twelve ratings of manifest abnormalities, a short formulation, and an *ICD* diagnosis. The detailed criteria for making the 5-point ratings and the definitions of the various abnormal phenomena have all been gathered together as a 'Manual' which may be consulted when difficult rating problems arise. The whole procedure takes from as little as ten minutes for a healthy individual with no medical complaints to as much as an hour for a patient with many symptoms.

This interview took us well over a year for its preliminary development and has naturally undergone a series of revisions aimed at increasing the reliability of the various ratings between different psychiatrists, and making the interview more suitable for the types of psychiatric disturbance that are commonly encountered in the community. When this had been done, we

TABLE 8.4. *Reported symptoms and manifest abnormalities rated at interview, showing reliability coefficients between interviewer and co-rater for each item*

Item	*Product moment correlation*	*Weighted kappa*
Reported symptoms		
Somatic symptoms	0·787	0·6733
Fatigue	0·805	0·7654
Sleep disturbance	0·981	0·8013
Irritability	0·841	0·6761
Lack of concentration	0·861	0·6745
Depression	0·914	0·8000
Anxiety and worry	0·830	0·6699
Phobias	0·789	0·7391
Obsessions and compulsions	0·837	0·6696
Depersonalization	0·870	0·7143
Manifest abnormalities		
Slow, lacking spontaneity	0·913	0·6000
Suspicious, defensive	0·858*	0·7368*
Histrionic	0·644*	0·4828*
Depressed	0·902	0·6646
Anxious, tense	0·773	0·6117
Elated, euphoric	0·981*	0·9362
Flattened, incongruous	0·804	0·7243
Depressed in thought content	0·766	0·6501
Excessive concern with bodily functions	0·829	0·7510
Thought disorder, delusions, misinterpretations	0·832	0·7113
Hallucinations	0·956	0·8789
Intellectual impairment	0·874	0·7753

* Based on twenty patients only.

embarked on a more extensive inter-rater reliability study with the interview in its present form. Each of the forty patients were interviewed by two psychiatrists, one of whom administered the interview while the other acted as co-rater. The interviews were arranged in two sets of twenty. In the first series of twenty interviews, five psychiatrists took part. Each psychiatrist carried out two interviews with each of his colleagues, once as interviewer and once as co-rater. The second series of twenty interviews was arranged in exactly the same way, except that there was one change of personnel.

Most of the standardized psychiatric interviews that have been developed for use with identified psychiatric patients have assessed reliability in terms of product moment correlations between observers, computed for each item scale. Kendell *et al.* (2) used a reliability coefficient called 'weighted kappa' in addition. This is a coefficient which takes account of the relative seriousness of the different types of disagreement that can occur between raters: it has a value of 1·00 if agreement is perfect, of zero if agreement is no better than that due to chance, and a negative value if agreement is less than chance.

TABLE 8.5. *Results of the three-way analysis of variance*

	SSq.	*d.f.*	*MSq.*
Source of variation			
Between patients	83·54	39	2·14*
Between items	74·47	19	3·92*
Between doctors	0·12	1	0·12 (N.S.)
Interactions			
Patients × items	259·37	741	0·35†
Patients × doctors	3·50	39	0·09 (N.S.)
Items × doctors	2·10	19	0·11 (N.S.)
Residual			
Patients × doctors × items	148·52	741	0·20
Total	571·61	1,599	

* Significant at 0·01 level. † Significant at 0·05 level.

We found that the product moment correlations for most of the 22 items were well above 0·8 and our values for weighted kappa were comparable with those reported by Kendell for the present state examination.

Kendell *et al.* in addition carried out a three-way analysis of variance on their data and we have also done the same with ours (Table 8.5).

For the purposes of this analysis, reported symptoms and manifest abnormalities have been grouped together simply as 'items'. When the mean squares for the main variables—doctors, patients, and items—were tested against the mean square for the residual it was found that while the mean squares for the patients and the items were significant at the 0·01 level, the mean square for doctors was not significant. This means that while there was significant variation between patients and between items, there was not significant variation between interviewer and co-rater. The significant interaction between patients and items merely indicates the tendency for different patients to have different symptoms. On the other hand, the small mean squares for the interactions between patients and doctors, and between items and doctors, are reassuring, since they indicate that the doctors' rating standards remained constant both from one item to the next and from one patient to the next.

Maxwell and Pilliner (3) have described a technique for using a three-way analysis of variance to estimate reliability between raters by comparing the between patient mean square with that for the interaction between patients and doctors. Using this method on the present data, the reliability coefficient between raters was calculated as +0·919.

We may therefore conclude that the reliability of the interview appears to be satisfactory and the interview seems to meet the requirements stated at the onset: it is economical of time, easy to administer, and flexible in that it can be adapted to the purposes of various types of inquiry. Perhaps most important, it has proved acceptable to a wide range of patients. So far it has been administered to over 400 general practice patients and to nearly as many individuals seen in a community screening programme; refusals and adverse responses have been very few. The clinicians who have administered the interview have found it appropriate for assessing and rating a wide range of psychiatric morbidity. A number of further studies will be required for complete evaluation of the interview: the high reliability found in the present study partly reflects the common background of the participants and the fairly extensive pre-training of the raters in pilot studies. Being only partly structured, the interview allows the psychiatrist who uses it a good deal of freedom; consequently it would not be reliable in inexperienced hands.

REFERENCES

1. Goldberg, D., Cooper, B., Eastward, M. R., Kedward, H. B., and Shepherd, M. (1970). 'A standardardized psychiatric interview for use in community surveys', *B. J. Prev. Med.* **24,** 18–23

2. Kendell, R. E., Everitt, B., Cooper, J. E., Sartorius, N., and David, M. E. (1968). 'The reliability of the "Present State Examination" ', *Soc. Psychiat.* **3,** 123–9.

3. Maxwell, A. E., and Pilliner, A. E. G. (1968). 'Deriving coefficients of reliability and agreement for ratings', *Brit. J. Math. Stat. Psychol.* **21,** 105–15.

Psychiatric morbidity and physical state in a general practice population

M. R. EASTWOOD

Institute of Psychiatry, University of London

Introduction

I should like to give a preliminary report on a study which has examined, in the community, the hypothesis that there is an association between physical and psychiatric disorder. There has long been an interest in this problem, since at least the time of the ancient Greek civilization, and it has usually been formulated in terms of the mind–body relationship. There is, however, no time to give a historical review of the literature on this occasion. I will restrict myself to mentioning some recent community studies which have dealt with this subject. Of particular interest have been the work by Downes and Simon (1) in Baltimore, Longaker and Godden (4) in the Stirling County study, Hare and Shaw (2) in their New Town study, and Shepherd, Cooper, Brown, and Kalton (6) in their survey of general practices in the London area. All of these showed a positive association between physical and psychiatric disorder.

However, these results have been difficult to evaluate because they have often been based upon the number of illnesses reported by the patients studied. This retrospective information was collected either in the form of self-assessments by patients or from medical records. The subjective bias and consulting habits of the patient and the standard of record-keeping may have influenced the results. Thus, should neurotic patients complain more in physical terms and have higher consultation rates than psychiatrically normal patients, they may give the impression of having more physical disease and actually, by their greater presence in the surgery, have more physical disease diagnosed.

Design and method

The study being described has the advantage that the psychiatric and physical assessments were made independent of consultation rates and as far as possible independent of symptoms. Also the physical and psychiatric diagnoses were made by different doctors. The study was so designed as to obtain a psychiatric sample from a random sample of the general population, to control it with an equal number of mentally normal people matched for age, sex, marital status, and social class, and to compare them for the presence of physical illness.

A health screening survey was held in a London borough, over a period of nine months, during which time approximately 1,500 people between the ages of 40 and 64 years, randomly

TABLE 8.6. *Twenty items from the Cornell Medical Index Health Questionnaire*

Number

121. Do you wear yourself out worrying about your health?
31. Are you often bothered by thumping of the heart?
32. Does your heart often race like mad?
80. Does pressure or pain in the head often make life miserable?
154. Do you wish you always had someone at your side to advise you?
148. Does your thinking get completely mixed up when you have to do things quickly?
181. Do you go to pieces if you do not constantly control yourself?
146. Do you get nervous and shaky when approached by a superior?
187. Do you often shake or tremble?
163. Does worrying continually get you down?
188. Are you constantly keyed up and jittery?
194. Do you often become suddenly scared for no good reason?
193. Do frightening thoughts keep coming back in your mind?
165. Does every little thing get on your nerves and wear you out?
108. Do you often get spells of complete exhaustion or fatigue?
113. Do you suffer from severe nervous exhaustion?
166. Are you considered a nervous person?
158. Do you usually feel unhappy and depressed?
161. Does life look entirely hopeless?
162. Do you often wish you were dead and away from it all?

selected from a general practice population, were screened. Psychiatric assessment and screening on a number of objective variables such as electrocardiography, blood pressure estimation, and serological values were carried out at a health centre. A fortnight later, when all these results had been collated, the patients were examined by their general practitioners.

Each patient completed a questionary pertaining to both physical and psychiatric disorder. The psychiatric items were 20 questions from the Cornell Medical Index Health Questionnaire. Statistical analysis of the data collected by Shepherd *et al.* (6) had shown that these were the items which best discriminated between general practice psychiatric patients and general practice normals. The 20 items and their numbers from the total CMI are shown in Table 8.6.

Results

Comparisons have been made between the scores on these 20 items for hospital out-patients, general practice patients, and the random sample used in this study. These are shown for the males in Fig. 8.1.

It can be seen that the curves, expressed as aggregate percentages, are different from each other, with those for the general practice psychiatric and normal patients being more similar to each other than either is to the hospital out-patients. The random sample contains many more nought scorers than the others and at the same time a few patients with very high scores. The curves for the females were similar, but there were less nought scorers. Prior to the screening it was decided that the most useful cut-off point for discrimination between psychiatric cases and normals would be five positive responses.

During the nine months of the survey, efforts were directed towards obtaining a psychiatric sample and a control group. Thus, anybody scoring on five or more items was considered to be a potential psychiatric case and those scoring nought to be potential controls. The mental states were assessed using the clinical schedule, just described by Dr Goldberg, both for the psychiatric group and the normal controls. The results are shown in Table 8.7. It can be seen from this table that the usual

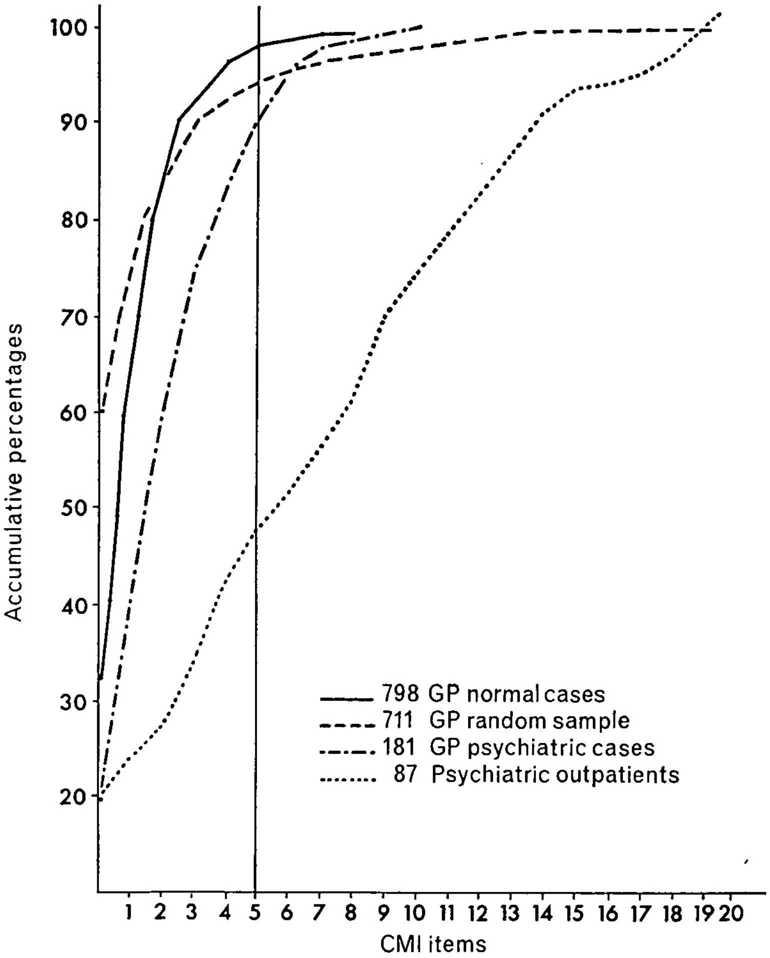


FIG. 8.1. Comparison of CMI scores in different populations, males

excess of females, commonly found in psychiatric surveys of the middle-aged, was found in this study.

Physical disorders

The physical disorders were operationally classified as being major or minor. A major condition was any physical disorder which threatened or was liable to shorten life, for example,

TABLE 8.7. *The screened population*

Population eligible for screening	=	2,075	
Population screened	=	1,461	(70·4%)
		711 males	
		750 females	
Potential index cases	=	58 males	(8·2%)
		184 females	(24·5%)
Potential control cases	=	426 males	(60·0%)
		267 females	(35·6%)
Total interviews for index cases	=	200	
Total interviews for control cases	=	169	
Final number of matched pairs	=	124	
		37 male pairs	
		87 female pairs	

ischemic heart disease. The minor conditions were the remainder of the physical disorders, for example, varicose veins. A large proportion of the major disease belonged to the cardiovascular and respiratory systems. It was not found feasible to measure duration of these physical disorders, although it was assumed that they would be largely chronic, since any acute condition would be liable to preclude attendance at the survey. The distribution of the major conditions by individual is shown in Table 8.8.

TABLE 8.8. *The distribution of major physical disease by individual in the index and control groups, by sex (percentage)*

Number of conditions by individual	*Male cases*		*Female cases*	
	Index	*Control*	*Index*	*Control*
0	51·4	67·7	35·6	50·6
1	27·0	33·3	35·6	35·6
2	16·2	—	24·1	11·5
3	5·4	—	3·5	2·3
4	—	—	1·2	—
Total	100·0	100·0	100·0	100·0
Total patients	37	37	87	87

Males: $P < 0{\cdot}02$. Females: $P < 0{\cdot}005$.

The average number of major physical conditions was significantly greater in the male and female index patients than in the controls. The correlation between psychiatric severity and the number of major conditions for the total psychiatric sample was positive and significant at the 5 per cent level.

Psychosomatic conditions

The physical conditions were placed into a 'psychosomatic' category, according to Sainsbury's (5) classification, whereby a condition had to be listed in the works of at least two out of six recent writers in the field. The distribution of major psychosomatic disorders is shown in Table 8.9. The average number of major psychosomatic conditions was significantly greater in the male and female index patients than in the controls.

TABLE 8.9. *The distribution of major psychosomatic disease by individual in the index and control groups, by sex (percentage)*

Number of conditions by individual	*Male cases*		*Female cases*	
	Index	*Control*	*Index*	*Control*
0	64·9	86·5	63·2	84·0
1	35·1	13·5	33·3	12·5
2	—	—	3·5	3·5
Total	100·0	100·0	100·0	100·0
Total patients	37	37	87	87

Males: $P < 0·05$. Females: $P < 0·01$.

Cardiovascular system

The only body system with numbers sufficient for comparison was the cardiovascular system. The electrocardiographic diagnoses were made using the Minnesota Coding which has been adopted by WHO. The distribution of cardiovascular disease is shown in Table 8.10. The psychiatric patients, both male and female, had significantly more major cardiovascular disease than the controls.

TABLE 8.10. *The distribution of major cardiovascular disease by individual in the index and control groups, by sex (percentage)*

	Male cases		*Female cases*	
Conditions by individual	*Index*	*Control*	*Index*	*Control*
No disease	62·0	86·5	69·0	84·0
Coronary heart disease	27·1	10·8	19·5	9·2
Hypertension	5·5	2·7	6·9	3·4
Hypertension and coronary heart disease	2·7	—	3·5	2·3
Congenital/rheumatic heart disease	—	—	1·1	1·1
Peripheral vascular disease	2·7	—	—	—
Total	100·0	100·0	100·0	100·0
Total patients	37	37	87	87

Males: $P < 0{\cdot}01$. Females: $P < 0{\cdot}025$.

Minor disorders

The distribution of minor physical disorders by individual is shown in Table 8.11. The average number of minor physical disorders was significantly greater in the male, but not in the female, index patients, compared with the controls. The correlation between psychiatric severity and the number of minor physical illnesses was not consistent. Thus it was positive for

TABLE 8.11. *The distribution of minor physical disease by individual in the index and control groups, by sex (percentage)*

Number of conditions by individual	*Male cases*		*Female cases*	
	Index	*Control*	*Index*	*Control*
0	27·0	54·1	41·4	46·0
1	37·9	32·4	46·0	31·0
2	32·4	10·8	10·4	16·1
3	2·7	2·7	1·1	2·3
4	—	—	1·1	4·6
Total	100·0	100·0	100·0	100·0
Total patients	37	37	87	87

Males: $P < 0{\cdot}02$. Females—no significant difference.

males and the older members of the total index sample but negative for females and the younger members of the total index sample. No significant differences were found for minor psychosomatic disorders.

Summary

In summary, there appears to be a positive association between physical and psychiatric disorder, in particular major physical disorder. Following on from the work of Hinkle and Wolff (3), it may be that there are individuals in the community who are susceptible to all forms of morbidity. It would be of great interest, proceeding from this work, for a prospective study to be undertaken, studying a cohort of patients, randomly chosen, to determine illness experience over a period of time.

REFERENCES

1. Downes, J., and Simon, L. (1953). 'Characteristics of psychoneurotic patients and their families as revealed in a general morbidity study', *Psychosom. Med.* **15**, 463.
2. Hare, E. H., and Shaw, G. K. (1965). *Mental Health on a New Housing Estate: A Comparative Study of the Health in Two Districts of Croydon* (London: Oxford University Press).
3. Hinkle, L. E., Christenson, W. N., Kane, F. D., Ostfeld, A., Thetford, W. N., and Wolff, H. G. (1958). 'An investigation of the relation between life experience, personality characteristics and general susceptibility to illness', *Psychosom. Med.* **20**, 278.
4. Longaker, W. A., and Godden, J. O. (1960). 'A comparison of organic and psychiatric symptoms in a small town', *Acta psychiat. scand.* **35**, 91.
5. Sainsbury, P. (1960). 'Psychosomatic disorders and neurosis in outpatients attending a general hospital', *J. Psychosom. Res.* **4**, 261.
6. Shepherd, M., Cooper, B., Brown, A. C., and Kalton, G. W. (1966). *Psychiatric Illness in General Practice* (London: Oxford University Press).

Psychiatric morbidity and social adjustment in a general practice population

BRIAN COOPER
M. R. EASTWOOD
JUDITH SYLPH

Institute of Psychiatry, University of London

Introduction

This paper outlines the method and preliminary findings of a current investigation of the social functioning of chronic neurotic patients in the community. Previous work has established that such patients constitute an important problem for the medical services. A survey of 46 practices in Greater London (10) found a one-year prevalence rate for all psychiatric disorders of 140 per thousand adults at risk. Just over half the declared cases were known to be chronic, as defined by the presence of continuous symptoms, disability, or psychopharmacological treatment for at least one year. Examination of a subsample (1) confirmed the case-identification in most instances and emphasized the serious nature of many cases not under specialist care.

At follow-up three years later, the clinical condition of the survey patients was reassessed by means of a number of techniques (7, 8). Of the original chronic group, only 44 per cent were now symptom-free, while 14 per cent were judged to have severe or moderately severe disorders. Comparison of a sample of the psychiatric patients with matched controls showed the former to have maintained consistently higher rates of consultation, specialist referral, and hospital admission (5).

These findings indicate the need for a systematic approach to the problems of chronic neurotic illness encountered in general practice. Measures which have been advocated include the more effective use of psychotropic drugs, the training of general practitioners in psychotherapy, and the establishment of firmer guidelines for specialist referral. All these lines of attack offer some prospects for the better treatment and management of neurotic illness within a strictly medical framework. So conspicuous, however, are the domestic and social problems presented by many neurotic patients that some observers have been led to urge a shift of focus from the individual patient to the nuclear family (6, 9) and from purely clinical treatment to a greater emphasis on social support (10, 11).

Experimental projects in general practice have confirmed the preponderance of psycho-social problems in the case-worker's load (2, 4). Developments of this kind are of great interest to the social psychiatrist but so far the necessary evaluative studies are lacking. There appears to be very little factual information on the nature and extent of social dysfunction among chronic neurotic patients, still less on the effects of social intervention in such cases. The present study is addressed to the first of these problems.

Design and method

The aims of the inquiry were, first, to test the hypothesis that clinically identified chronic neurotic patients in the community are characterized by impaired social functioning; secondly, to examine the nature and frequency of occurrence of any characteristic social difficulties. To this end, it was planned to compare a representative sample of chronic neurotic patients with a matched control group of mentally normal individuals drawn from the same general practice population. This basically simple design gave rise to a number of problems of method which may be enumerated as follows:

I. DEFINING THE PATIENT POPULATION

The patients were selected from the NHS lists of general practitioners in Croydon and south-east London who had agreed to co-operate in the research project. Each doctor kept a special record for one month of all surgery consultations, including the patient's name, age, marital status, occupation, and diagnosis or presenting symptoms. For patients who were not gainfully occupied, their own previous occupation or that of the chief breadwinner was recorded, as appropriate. Individual practices were taken in sequence so that the time-sample of consultations was drawn in a different month of the year for each of the doctors who took part.

2. SELECTING THE INDEX GROUP

Each practice list was collected week by week and examined for psychiatric diagnosis. All such cases were discussed briefly with the recording doctor, and, in addition, the medical records were scrutinized. Any new or recent episodes were discarded and only those patients with evidence of neurotic ill-health of at least one year's duration were accepted for the sample. Patients reported to be only mild or border-line cases were likewise excluded, since the object of the screening procedure was to obtain two clearly contrasting patient groups. At the other end of the spectrum, clearly psychotic patients were also excluded because of the special nature of the social problems they present.

Each patient accepted as a potential index case was then approached by his family doctor and asked to attend the surgery by appointment for an interview with one of the two research psychiatrists. Those who complied were given a standard psychiatric interview, details of which have been described by Dr Goldberg (see above, and Goldberg *et al.* (3)). As a result, each patient received a series of individual symptom ratings, a psychiatric diagnosis and an over-all severity score. Any patients not confirmed at interview as clinical psychiatric cases were dropped from the sample at this stage; the remainder were asked to agree to a home interview with one of the social investigators.

3. SELECTING THE CONTROL GROUP

Potential control patients were selected from the same surgery attendance lists as the index group. Individual pairs were matched from the lists for sex, age-group, marital status, occupational status, and social class. Immigrants and persons living alone were matched with like individuals. So far as possible, each index patient was matched with at least two potential controls who could be approached in turn. Since on average each practice provided about 300 patients from which to match a dozen index cases, it was possible in most instances to find suitable controls within the practice. Where matching proved difficult, controls were selected from one of the other survey practices.

The potential control patients thus selected were approached and interviewed in exactly the same way as the index patients. Any who at clinical interview showed evidence of psychiatric disturbance were rejected as unsuitable; in all such instances an alternative control was chosen. The social interview also followed the same lines as for the index patients, the social investigators being unaware which patients were index cases and which controls.

This method has the advantage that information on all the important demographic and social variables can be obtained economically for large numbers of people. The chief disadvantage is that since the control group is drawn exclusively from surgery attenders, it is likely to carry an excess of general morbidity as compared with the general population. Inasmuch as general morbidity is associated with social handicaps, such a bias would tend to reduce the differences in social adjustment to be found between index and control groups, and thus to produce a negative result.

4. MEASUREMENT OF SOCIAL ADJUSTMENT

Information on social conditions, functioning, and attitudes was collected by means of a standardized, semi-structured interview carried out in the patient's home. In all cases where the patient was living in a family group, a key relative—usually spouse or parent—was also present during the interview. Patients whose key relatives were unwilling or unable to participate were discarded from the study.

Since the social and domestic circumstances of the patients varied widely, a flexible interview technique was essential. The standard interview which was employed—and which is still in process of development—contains sections on housing conditions, employment, family income level, leisure activities and social contacts, marital adjustment, child management, and personal interaction with other members of the household, relatives, neighbours, and workmates. Each item is rated on a 4-point scale ranging from 0 ('satisfactory, no difficulties') to 3 ('severe difficulties or dissatisfaction') and assigned to one of the following categories:

1. *Material conditions*

Comprising relatively objective factual data on housing conditions, family income, and situational handicaps (including chronic physical illness and disability).

2. *Management and competence*

As shown by the patient's ability to conduct his social affairs and personal relationships, e.g., standards of marital adjustment and interaction with relatives and neighbours.

3. *Satisfaction*

As judged from the patient's own expressed attitudes to his living conditions, occupation, and social role, e.g., as housewife, parent, retired person, etc.

It was hoped that this tripartite division would allow social dysfunction to be defined with some precision and would, moreover, provide guidance as to the most appropriate forms of social intervention and support. On the evidence of a pilot study of twenty matched pairs (11), the three categories appeared to provide a useful basis for analysis of the findings. The framework of the interview is set out in Table 8.12.

Clearly, the individual items could not be applied uniformly to all patients, so that direct comparison of the findings for index and control groups depended upon the success of the matching procedure. In practice, no control patients were accepted who could not be rated on the same items as the corresponding index patients.

The standardized social interview appears to be of good overall reliability in this type of setting. Joint interviews on a series of 17 general practice patients with a fairly wide range of social problems demonstrated good agreement between two investigators trained to administer the schedule (excluding 0 ratings, $r = +0{\cdot}91$). Agreement of the same order was found for about half the individual items, the rest having too few positive ratings

TABLE 8.12. *Classification of social interview ratings*

Material conditions	*Management and competence*	*Patient's satisfaction*
Housing conditions	Household care	with house and district
—	Occupational stability	with employment with social role
Household income	Management of money	with financial state
—	Marital adjustment	—
Situational handicaps to:	Standard (and extent) of:	
Leisure and social activities		with leisure and social activities
Interaction with relatives		
Interaction with workmates		
Interaction with neighbours		
Child management		with parental role

to be testable; from this point of view more extensive field studies will be required. A detailed account of the development of the social interview is in preparation.

Preliminary findings

At the time of writing (July 1969), interviewing is still in progress and so far some 80 matched pairs have been rated in six practices. The results here presented concern only the first 60 matched pairs. The response to the inquiry by the neurotic patients has been encouraging: of 101 classed as potential index cases in the first five practices, 92 (91·1 per cent) attended for clinical interview. All these patients manifested neurotic disturbance in some degree, but only 74 fulfilled the operational criteria, the remaining 18 not having had symptoms continuously throughout the preceding twelve months. Of the 74 confirmed chronic patients, 69 (93·2 per cent) were later interviewed at home by the social investigators. It seems probable, therefore, that the sample thus obtained is broadly representative of chronic neurotic illness in the practices concerned.

Selection of the control patients posed more difficult problems, partly because in this group the patient's response was not as good; partly because 'normals' proved less reliably identifiable from the records than 'cases'. Of 79 patients selected as potential controls in the first five practices, 14 (17·7 per cent) either refused to take part or failed to attend the surgery for interview. Thirteen of the 65 patients who did attend (20·0 per cent) were considered unsuitable as controls because they were suffering from neurotic conditions usually mild in nature, which had not been diagnosed by their own doctors. The first of these groups represents a source of bias which must be considered when the findings of the study are interpreted. The second stems directly from the research design, the object of which was to isolate two clearly contrasting end-groups from the spectrum of neurotic morbidity which exists in the general population.

TABLE 8.13. *Psychiatric interview ratings: aggregate scores for sixty matched pairs*

	Males (n = 14)		*Females* (n = 46)	
	Index	*Control*	*Index*	*Control*
Reported symptoms				
Somatic symptoms	24	2	66	14
Fatigue	22	3	70	20
Insomnia	30	8	71	20
Irritability	19	1	52	12
Loss of concentration	17	5	52	6
Despondency	21	2	74	6
Depressive thoughts	20	1	55	1
Anxiety	25	2	83	17
Phobias	11	1	54	18
Obsessions	17	1	40	11
Depersonalization	0	0	16	1
State at interview				
Hypochondriacal	13	1	38	13
Histrionic	5	1	24	4
Anxious	24	1	81	9
Depressed	21	0	54	1
Retarded	5	0	6	0
Elated	0	0	0	1
Flat or incongruous mood	0	1	2	0
Delusional state	0	0	8	0
Intellect impaired	2	1	4	1

The extent to which this aim was successful in practice is illustrated by Table 8.13 in which the clinical item scores for the index and control groups are compared.

The table demonstrates, first, that the two groups were clearly differentiated for both sexes in respect of psychiatric morbidity; secondly, that the distribution of symptoms and signs was in conformity with that predominance of neurotic depressive reactions and anxiety states which has been reported in most general practice and community surveys.

TABLE 8.14. *Distribution of social interview ratings for the index and control groups*

	Males (n = 14)		*Females* (n = 46)	
Social interview item score	*Index*	*Control*	*Index*	*Control*
0	115	174	385	559
1	49	44	217	182
2	40	13	123	39
3	29	10	77	19

In Table 8.14 the distribution of social interview ratings is compared for the two groups. The number of 0 ('satisfactory') ratings is higher for both sexes in the control group, while ratings of 2 and 3, signifying relatively serious dysfunction, are markedly higher among the index patients.

TABLE 8.15. *Comparison of main categories of social dysfunction: aggregate scores for the index and control groups*

	Males (n = 14)		*Females* (n = 46)	
Categories of social dysfunction	*Index*	*Control*	*Index*	*Control*
Material deficiencies	71	54	221	157
Social management problems	76	30	267	87
Dissatisfaction	69	16	206	73

Finally, in Table 8.15 these findings have been broken down into the three main social adjustment categories of material conditions, social management and satisfaction. While the same trend is apparent for both sexes in each of these categories, it can be seen that the main differences lie in the management and satisfaction ratings. The differences in material conditions are significant for the female patients ($t = 2{\cdot}59$; $n = 46$; $p < 0{\cdot}01$), largely due to the presence in the index group of a number of women suffering from pronounced social isolation; in general, however, they are much less marked than those in the other two main categories.

Discussion

The preliminary findings outlined above provide support for the hypothesis that in a general practice population, neurotic illness, clinically defined, will be found to be positively associated with social dysfunction. Some qualification must be made, however, as to the representative nature of the present findings. While the method of investigation here described has powerful advantages for comparative matching, it cannot overcome two other methodological problems common to this type of inquiry: that of non-response or refusal to co-operate on the part of patients and that of the non-random selection of practices.

The problem of bias due to non-response is endemic in field surveys and is to some extent insuperable. In a general practice context, it can be reduced with the help of certain doctors who have an exceptional ability for obtaining the co-operation of their patients. One such man took part in the present study and it is noteworthy that although the response rate in his practice was indeed very high, the differences between his index and control patients were closely similar to those found for the sample as a whole.

The effect of non-response could not be accurately gauged, but in the index group the chief result appeared to be the under-representation of certain types of social problems; in particular several cases of marital conflict were lost because either husband or wife was unwilling to agree to a home interview. Among the control patients, where it occurred more frequently, non-response may have been responsible for a degree of bias in the findings. One could, for example, postulate that control patients who failed to attend for interview had an above-average number of social problems which they were reluctant to disclose. In general, however, our impression was that non-response was linked rather with a full programme of activities and the absence of any pressing health or social problems. The very good response of the neurotic patients may, indeed, have been attributable to their need to discuss their health worries and personal difficulties.

Our experience in mounting this study confirmed earlier findings (10) regarding the impossibility of procuring a random sample of practices for collaborative research. How far the findings can be extrapolated to a wider population remains, therefore, an open question. The practices included so far vary widely in their social structure and are thought to provide a fairly good cross-section of the local population. To the extent that this study represents the first phase of an operational research programme, we are not immediately concerned with its relevance for other areas, although clearly it would be of some interest to replicate the survey in a number of widely contrasting urban and rural settings.

The next phase of our programme will consist of a one-year follow-up of the index patients from this study, designed to measure any changes in their clinical state and social adjustment and to test for correlation between these variables. We shall also be concerned with the patients' use of the available medical and social services. So far, the findings indicate that these services are making relatively little impact on the type of ill-health under investigation: at the time of interview only 12 per cent of the index patients were under psychiatric supervision while less than one-third had been in touch with any psychiatric or social agency during the preceding twelve months. The chronic nature of the patients' symptoms and the high prevalence among them of social coping difficulties point to the need for long-term case-work by trained medical social workers. It is hoped that the findings of the present study will serve as a baseline for future evaluative studies in this field.

REFERENCES

1. Cooper, B. (1965). 'A study of 100 chronic psychiatric patients identified in general practice', *Brit. J. Psychiat.* **111**, 595.

2. Forman, J. A. S., and Fairbairn, E. M. (1968). *Social Casework in General Practice* (Oxford University Press for Nuffield Provincial Hospitals Trust).

3. Goldberg, D., Cooper, B., Eastwood, M. R., Kedward, H. B., and Shepherd, M. (1970). 'A standardised psychiatric interview for use in community surveys', *Brit. J. prev. soc. med.* **24**, 18.

4. Goldberg, E. M., Neill, J., Speak, B. M., and Faulkner, H. C. (1968). 'Social work in general practice', *Lancet*, **ii**, 552.

5. Harvey-Smith, E., and Cooper, B. (1970). 'Patterns of neurotic illness in the community', *J. Roy. Coll. gen. Practit.* **19**, 132–9.

6. Howells, J. G. (1963). *Family Psychiatry* (London and Edinburgh: Oliver and Boyd).

7. Kedward, H. B. (1969). 'The outcome of neurotic illness in the community', *Soc. Psychiat.* **4**, 1.

8. Kedward, H. B., and Cooper, B. (1966). 'Neurotic disorders in urban practice: a three-year follow-up', *J. Roy. Coll. gen. Practit.* **12**, 148.

9. Ryle, A. (1967). 'Neurosis in the ordinary family', *Mind and Medicine Monograph No. 15* (London: Tavistock Publications).

10. Shepherd, M., Cooper, B., Brown, A. C., and Kalton, G. (1966). *Psychiatric Illness in General Practice* (London: Oxford University Press).

11. Sylph, J., Kedward, H. B., and Eastwood, M. R. (1969). 'Chronic neurotic patients in general practice', *J. Roy. Coll. gen. Practit.* **17**, 162.

9

SOME CURRENT RESEARCH STUDIES IN PSYCHIATRIC EPIDEMIOLOGY

Social class and stressful events

BARBARA S. DOHRENWEND
The City College of the City University of New York, U.S.A.

The work I shall report is part of the program of research that Bruce Dohrenwend described in another session of this symposium. I will pick up the point he made concerning the need to consider possible transient responses to stressful events in attempting to explain variations in reported rates of psychological disorder.

We start from the empirical generalization, drawn from epidemiological studies of untreated as well as treated psychological disorder, that the highest rates of symptoms are found in the lowest social class (3). One possible explanation of this excess of symptoms might look like this:

Lower class → High rate of stressful events → High rate of transient psychological symptoms.

Let us consider the second part of this relationship first. There is ample evidence from studies of events ranging from bombing of civilian populations (6) to removal of slum dwellers from their homes to make way for urban renewal (5) that unusual stressful events do produce symptoms of distress in a large proportion of those exposed to them. Moreover, in a recent cross-national convergence, two investigations demonstrated that a variety of more ordinary changes in life circumstances was associated with symptoms of psychological disorder. In Britain, Brown and Birley (1) showed that in the three-week period preceding the onset of an episode of schizophrenia, patients had experienced significantly more 'events which on common-sense grounds are likely to produce emotional disturbance' (p. 204) than had a healthy control group. In a community study in Washington Heights, a section of New York City, Bruce Dohrenwend and I showed that a variety of unfavourable life changes (such as illness in the family or loss of a job) were associated with increases, while favourable life changes (such as a birth in the family or a promotion) were associated with decreases in scores on the Midtown 22-item inventory of neuropsychiatric symptoms (4). Thus earlier work has provided support for the second part of the relationship described above, suggesting that a group which experiences a relatively high rate of stressful events might, as a consequence, exhibit a relatively high rate of psychological symptoms. For this reason, we have turned our attention for the time to the first part of the relationship, that is, to the question of whether there is an excess of stressful events in the lower class.

How shall we define stressful events for this purpose? In developing our definition, we have had two considerations in mind. First, we want to be sure that the events we study are not stressful primarily as a consequence of antecedent psychological disorder. One way in which we deal with this problem is to define events as stressful in objective terms rather than asking the respondent what he considered stressful, reasoning that the fact that a person considered an event stressful might well say more about his condition at the time the event impinged on him than about the nature of the event (2).

The second major consideration that we took into account in defining stressful events was that we did not want to prejudge what was a trivial and what was an important event. Thus, the general definition that we have used is any event which, for better or for worse, potentially changes or interrupts the subject's usual activities. The consequences of this inclusive definition are illustrated in Table 9.1 where you will see, among others, the seemingly trivial event of 'acquiring a cat'. We would not rule out the possibility that even this event might be disturbing, particularly if the animal had not been housebroken.

Design of a pilot study

As the first step toward testing our hypothesis concerning the relationship between social class, stressful events, and psychological symptoms, we conducted a small pilot study of the relationship of social class to rate and quality of objective change-producing events. This interview survey was conducted with 97 community residents who were heads of households in two census tracts in Washington Heights, New York. These two

TABLE 9.1. *Examples of events elicited by checklist interview question classified according to direction of status change*

	Direction of status change	
Activity area	*Gain*	*Loss*
Marriage and family	Birth of child	Divorce
Health	Health improved after illness	Broke a finger
Education	Started nursing school	Funked out of college
Occupation	Promotion	Laid off
Migration	Moved to better neighbourhood	Evicted from apartment
Recreation	Took a vacation	Quit stamp-collecting
Miscellaneous social changes	Made new friends	Daughter got into fight with girls in the neighbourhood
Miscellaneous material changes	Acquired a cat	Car stolen

tracts were chosen for ethnic and income heterogeneity, in order to obtain a sampling of events experienced by a variety of population groups.

The sample of respondents was divided according to income, with the lowest feasible cutting point chosen in order to separate lower class from other respondents as clearly as possible. The divison at $4,000 of family income per annum placed 20 of the respondents in the lower-income group.

The analysis we report includes only respondents who were below retirement age, since the retired group differed from younger respondents in number and types of events reported and were, at the same time, over-represented in the lower-income group. While women and Negroes were also over-represented in the lower-income group, the class differences that we shall report cannot be explained as simply sex or race differences. The possibility of complex effects due to the interaction of sex or race and income cannot be dismissed, but tests for such effects will require a larger sample than we have at present.

Number of events

In the interview, the first question designed to elicit change-producing events was, 'what was the last major event in your life that, for better or for worse, changed or interrupted your usual activities?' Following this question, the respondent was asked to select all events that he had experienced in the last two years from a checklist which, in its final form, covered the activity areas in the right column of Table 9.1, and included activities of family members and significant others as well as the respondent's activities. Our aim was to get from each respondent an exhaustive report of events that had changed his life in any way, temporarily or permanently, within a specified time.

TABLE 9.2. *Comparison of number of change-producing events during two years preceding interview in two income groups*

	Income	
Number of events	*Less than $4,000*	*$4,000 or more*
Mean	2·08	1·79
SD	1·80	1·50
Number or respondents*	13	58

$t = 0{\cdot}35$. $df = 69$. $p > 0{\cdot}05$.

* Excluding 18 respondents at or above retirement age of 65 and 8 respondents who failed to provide information about income or about dates of events.

As Table 9.2 shows, there is no evidence that the lower income group had experienced a significantly higher rate of such events than persons with higher incomes in the two-year period preceding the interview. This result is consistent with our inference from a review of the extensive literature on class differences (3) that the frequency of objective change-producing events probably does not vary with social class.

Favourableness of events

For further analysis, the events reported by respondents were classified as to whether, in terms of culturally defined norms, they represented a gain or a loss in status or were ambiguous.

This classification is illustrated in Table 9.1. Our prediction, based on our review of the literature on class differences, was that the proportion of events involving loss of status would be greater in the lower than in other social classes.

When a respondent reported more than one event for the two years prior to the interview, he was asked which he considered most important. Table 9.3 shows that, taking a single event for each respondent, either the only one reported or the most important one, the evidence indicates that the events experienced by the lower-income group were more likely than events experienced by higher-income groups to involve loss of status. Most often these lower-class losses consisted of illness or death in the family. The class difference in relative frequency of status gain and loss is found also among events not considered most important.

TABLE 9.3. *Proportions of respondents in two income groups for whom most important event in last two years involved loss rather than gain in status**

	Income	
	Less than $4,000	*$4,000 or more*
Percentage	88·8	47·6
Number of respondents†	9	42

Fisher exact probability = 0·03.

* Two independent coders agreed for 91 per cent of events on classification into categories: gain in status, loss in status, ambiguous whether gain or loss.

† Excluding 8 ambiguous cases: n is also reduced from that in Table 9.2 because 11 respondents reported no event in two-year period and one respondent did not indicate which of several events was most important.

Respondent's responsibility for events

In our analysis of class differences we are also concerned about an issue dealt with both by Brown and Birley (1) and in our earlier Washington Heights study (3), to wit: are the reported events in part or in whole the responsibility of the person affected? In order to answer this question the events reported by our respondents were classified into the categories shown in Table 9.4.

TABLE 9.4. *Relation of income to respondent's responsibility for most important change-producing event in two years preceding interview (percentage)*

*Category of responsibility**	*Income* *Less than $4,000*	 *$4,000 or more*
Respondent probably entirely responsible for event	0·00	12·8
Respondent probably shared responsibility for event with a peer	16·7	23·4
Respondent probably influenced event, largely controlled by a person in superordinate status	8·3	25·5
Sum of above: respondent probably had some degree of responsibility for event	25·0	61·7
Respondent probably had no responsibility for event	75·0	38·3
Number of respondents†	12	47

Fisher exact probability = 0·02.

* Two independent coders agreed for 74 per cent of events on classification into four categories and for 91 per cent of events on classification into two categories: some responsibility, no responsibility.

† Less than n in Table 9.2 because 11 respondents reported no event in two-year period and one respondent did not indicate which of several events was most important.

The distinction with which we are primarily concerned is between events for which the respondent could be expected to have had some degree of responsibility, categories one to three in Table 9.4, and events for which he could not be expected to have had any responsibility, category four. The inference that an event was the cause of psychological symptoms rather than part of a larger syndrome indicating underlying psychological disorder would clearly be more defensible when the respondent had had no responsibility for the event.

Table 9.4 shows that the lower-income group was more likely than higher-income groups to report events for which they probably had no responsibility. Furthermore, of the nine events in Table 9.3 involving loss of status and reported by lower-income respondents, only one was classified as partially the responsibility of the respondent. Thus, these results suggest that the relatively large number of unfavourable change-producing events suffered by the lower class could not be attributed to their making trouble for themselves.

In another sense, however, there may be an issue of responsibility in relation to the class difference in favourableness of events. Whereas only a quarter of all events in Table 9.3 involving loss of status were classified as probably the responsibility of the respondent, twenty (95 per cent) of events involving status gain were deemed the respondent's responsibility in some degree. Thus, the question arises whether the general lack of status gains in the lives of the lower-class respondents can be attributed to their failure to act on their own behalf to achieve favourable changes. We cannot, unfortunately, answer this question with the present data.

Although a number of questions are left unanswered by this pilot study of the relation between social class and stressful life events, we find the results promising. Specifically, we think they indicate that this line of inquiry may yield an explanation of at least some portion of the excess of psychological symptoms consistently found by epidemiologists in persons in the lowest social class.

REFERENCES

1. Brown, G. W., and Birley, J. L. T. (1968). 'Crises and life changes and the onset of schizophrenia', *Journal of Health and Social Behaviour*, **9**, 203–14.

2. Cumming, J., and Cumming, E. (1962). *Ego and Milieu* (New York: Atherton Press).

3. Dohrenwend, B. P., and Dohrenwend, B. S. (1969). *Social Status and Psychological Disorder: A Causal Inquiry* (New York: J. Wiley).

4. Dohrenwend, B. S., and Dohrenwend, B. P. (1966). 'Stress situations, birth order, and psychological symptoms', *Journal of Abnormal Psychology*, **71**, 215–23.

5. Fried, M. (1963). 'Grieving for a lost home', in *The Urban Condition*, pp. 151–71, ed. Duhl, L. J. (New York: Basic Books).

6. Janis, I. L. (1951). *Air War and Emotional Stress* (New York: McGraw-Hill).

Social precipitants of severe psychiatric disorders

GEORGE W. BROWN and J. L. T. BIRLEY
Bedford College, London, and the Institute of Psychiatry, University of London

Opinion is still sharply divided about the role of life changes and crises as precipitants of severe psychiatric disorders. Both research and criticism have been almost entirely limited to common-sense judgements about individual cases. Beck (2), for instance, in a recent review of research on depression, considers as unconvincing certain events quoted in an article by Hamilton and White (5). These were: the patient's being left alone for prolonged periods while his wife went to look after their sick daughter; another patient's being put in charge of a work programme beyond his capacity; and, for a third case, learning that the primary tuberculosis which he had had for nine years was bilateral.

Such arguments about severity will clearly get us nowhere. What is needed to begin with is a standardized approach to ascertain for a series of patients the temporal link between certain events in certain onsets. The accurate dating of onset and events is of critical importance. But such research will not be enough. Evidence cannot be convincing without the solution of some troublesome methodological problems. Two are particularly important. Firstly, patients and relatives may, in recalling the past, exaggerate the significance of events as a means of trying to come to terms with the illness. Bartlett (1) has called this tendency 'effort after meaning' and, if not controlled, it will tend to increase the number of so-called precipitating events. Secondly, it is quite clear from published reports that many 'precipitating events' could simply be due to the insidious onset of the illness itself. For example, the patient just quoted may have found the work beyond him because of his developing illness.

In a recent publication we have described ways in which we have attempted to meet such difficulties (4). For example, in questioning about possible precipitating events we avoid asking or discussing what may have been found distressing or exciting, and simply check through a standard list of events which on common-sense grounds many would expect to find emotionally disturbing. Exactly the same list of events is covered with everybody; and each possible event is defined in such a way as to make it unlikely that it would be included simply because the informant is, in retrospect, exaggerating its significance. For example, events occurring to so-called friends are not included, as it is easy to exaggerate the degree of closeness between the patient and his acquaintances. Second, we include in the main analysis only events which on logical grounds would be most unlikely to have been brought about by any developing disorder. For this, criteria were set up before the main research began. Most of the events, we believe, are not subject to the patient's control. Losing a job because a whole firm closed down would be included, for example, but not losing a job for more personal reasons. Such rules mean, of course, that the research is likely to establish only a conservative estimate of the role of social events in precipitating mental disorder.

We have completed one study of fifty schizophrenic patients and have started collecting material on a much larger and more extensive study of depressive conditions.[1] The schizophrenic patients all had major and clear-cut changes in their condition in the three months before admission and formed about 40 per cent of all admissions with a diagnosis of schizophrenia. They showed a marked excess of important life changes and crises in the three-week period before onset in comparison with the three previous three-week periods. As already mentioned, all events were judged to be most unlikely to have been brought about by any unnoticed and insidious development of the disorder. Forty-six per cent had at least one such event in the three-week period immediately before onset, in comparison with 14 per cent, 8 per cent, and 14 per cent respectively in the three prior three-week periods. The rate of such events in a comparable group of normal subjects was almost the same in all

---

[1] This study is being carried out with Mrs Freda Sklair and Miss Jennifer Frankland of Bedford College, London, and Dr John Copeland of the Institute of Psychiatry.

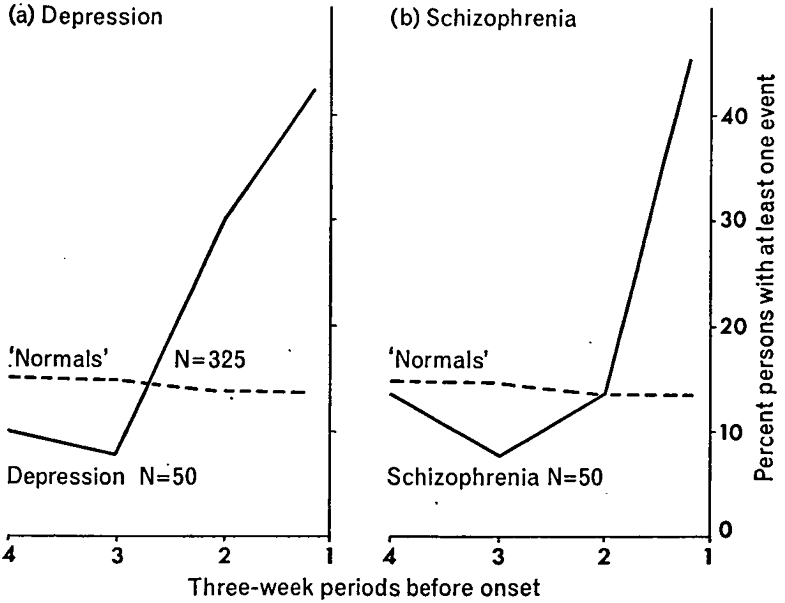


FIG. 9.1. Percentage of persons with at least one event in the four three-week periods before onset (patients) or interview (general population)

four three-week periods before interview, averaging 14 per cent, a rate very similar to the patient group outside the three-week period nearest onset (Fig. 9.1).

In our current study of depressed patients we have taken any who have shown marked deterioration in the twelve months before admission or (if not admitted) attendance at out-patients. Relatively few patients have been excluded on clinical grounds and patients therefore cover reasonably well the wide range of depressive disorders treated by psychiatrists. Results concerning events in the twelve weeks before onset for the first fifty cases are similar to those for schizophrenic patients. Forty-two per cent had an event in the three-week period before onset of the first psychiatric symptoms. However, they also had an increased rate of events in the second three-week period—30 per cent had at least one event in the fourth to sixth week before onset—so that a somewhat longer period before onset may be critical for depressed patients. Thirty-one of these fifty depressed patients also had a second major worsening of their condition, on average seventeen weeks after the first change. Fifty per cent of these patients had an event in the three-week period immediately before this second change in their condition. There is evidence, therefore, that life changes and crises often play an important role in all the stages of a developing depressive disorder.

We believe that we have made a reasonably strong case for the importance of life changes and crises in precipitating onset and relapse in schizophrenic and depressive conditions. A whole series of further questions, of course, now demand answers.

We are particularly concerned with:

(*a*) The characteristics of the events themselves—for example, are unexpected events more likely to lead to disorder?

(*b*) Is there anything about the person's prior experience, preparation, immediate reaction or his coping with the consequences, and implications of the event which makes breakdown more likely?

(*c*) What about situational factors? Do, for example, other events in the year or long-term difficulties play a role? Does support received from the environment, including that from family, friends, and work-group, or the person's recognition of alternative sources of satisfaction, make breakdown less likely?

(*d*) Finally, does the patient's diagnostic group or pattern of symptoms relate to such factors? Do schizophrenic and depressive disorders, for example, tend to be precipitated by different kinds of event?

So far we have analysed only material for the less-detailed study of schizophrenic patients, but results suggest that fairly clear answers can be obtained to such questions. For example:

(*a*) There is no difference in symptomatology, as judged by frequency of individual symptoms, between those patients who had and those who had not experienced a precipitating event. Depressive delusions and preoccupation with death, but not other depressive symptoms, were expressed rather more frequently by those patients who had experienced a precipitating event. Among so many comparisons, however, these might well be due to chance (3).

(*b*) Situational factors do apparently play an important role. For example, schizophrenic patients were much more likely to have had an event in the three weeks before onset if they had been living in homes which we had independently characterized as at all tense, suggesting some kind of additive process (4).

In our current study of depressed patients, we are examining such problems in much greater detail—interviews with the patient and relatives take in all a total of about six hours. We also plan to include comparable material from a large series randomly selected from the general population.

REFERENCES

1. BARTLETT, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology* (Cambridge University Press).
2. BECK, A. (1967). *Depression: Clinical Experimental Structural Aspects* (New York: Harper and Row).
3. BIRLEY, J. L. T., and BROWN, G. W. (1970). 'Crises and life changes preceding the onset and relapse of acute schizophrenia: clinical aspects', *Brit. J. Psychiat.* **116**, 327–33.
4. BROWN, G. W., and BIRLEY, J. L. T. (1968). 'Crises and life changes and the onset of schizophrenia', *Journal of Health and Social Behaviour*, **9**, 203–14.
5. HAMILTON, M., and WHITE, J. (1959). 'Clinical syndromes in depressive states', *J. ment. Sci.* **105**, 485–98.

Bereavement as a precipitating event in mental illness

ZENA STEIN
Associate Professor of Epidemiology and Director
New York State Department of Mental Hygiene,
Epidemiology Research Unit, Columbia University, U.S.A.

and

MERVYN W. SUSSER
Professor and Chairman, Division of Epidemiology,
Columbia University, U.S.A.

In this paper we examine the relationship between widowhood and mental illness, as a means of approaching the general question of social and psychological causes of mental illness. A demonstrated effect of widowhood on mental illness can reasonably be taken as a special case of the social and psychological effects of bereavement.

Three sets of data drawn from a single community are presented in this paper. One set is derived from a descriptive study of the incidence of mental illness; these data strengthen the reported associations of widowhood with psychiatric care for mental illness and argue against a loss of kin support as a reason for the association. A second set of data derives from a prevalence study of chronic disability arising from mental disorder; these data indicate that there is no association between the widowed state and such chronic disability. A third set of data is derived from an analytical study aimed at testing a possible causal effect of widowhood on mental illness; these data bear on the time interval between bereavement and first entry into psychiatric care, and support the hypothesis that the transition from the married state to widowhood precipitates entry into care for mental illness.

Widowhood and inception of psychiatric care

From 1959 to 1963 a register was maintained of all adult cases from Salford, England (population 155,090 in the 1961 census) that came into any form of special psychiatric care. Every fresh referral to an agency providing psychiatric care was treated as an episode and among them the 1,945 persons registered as seeking psychiatric care for the first time in their lives were classed as inceptions. Both the method of this incidence study and the raw data on which the analysis is based are recorded in detail elsewhere (1, 4). Marital state, age, sex, social class, and diagnosis were recorded for all episodes and the distributions of these attributes among patients were computed as rates against the 1961 census.

TABLE 9.5. *All mental illness: inceptions by sex and marital state*

	Average annual rate per 100,000 in 1961 census, Salford, England (1959–63)*			
	Single	*Married*	*Widowed*	*Divorced*
Men	493	214	586	550
Women	505	321	534	758

* Standardized against the total Salford population, for men and women separately.

Table 9.5 shows the rates of inception of all mental illness by sex and marital state at the time of inception, standardized for age against the total Salford population of 1961. The rates of inception were much higher for the widowed than for the married.

Table 9.6 shows age-standardized rates of inception for selected diagnoses.

The widowed had higher rates than the married for almost all diagnoses, notably for depressive psychosis among women, for addictions (most often alcoholic psychosis) among men, and for organic and senile dementias among both sexes. Only for psychoneuroses were the rates for the widowed lower than for the married among both sexes.

TABLE 9.6. *Inceptions of selected diagnoses by marital state* in each sex*

		Average annual rate† per 100,000 in 1961 census, Salford, England (1959–63)		
		Single	*Married*	*Widowed*
Schizophrenia	Men	101	14	32
	Women	60	23	23
Depressive psychosis	Men	98	56	85
	Women	152	118	210
Psychoneurosis	Men	73	49	45
	Women	91	77	27
Addiction	Men	5	7	49
	Women	5	3	1
Psychopathy	Men	46	18	24
	Women	41	11	27
Organic	Men	53	12	27
	Women	17	15	39
Senile	Men	33	16	26
	Women	57	34	49

* The divorced were considered too few in number to permit the refinement of analysis required by this table.

† Standardized against the total Salford population, for men and women separately.

Table 9.7 shows inceptions for all mental illness by marital state, sex, and age. In every category (save the 15–19 age-group) the widowed have higher rates than the married, the differences being most marked in the 20–39 age-group. This finding among the young widowed is reminiscent of the high rates for mortality among the young widowed observed by Kraus and Lilienfeld (2) and the high rates for suicide among the young widowed observed by MacMahon and Pugh (3).

Widowhood and prevalence of chronic disability associated with mental disorders

In the same community in which the register of inceptions was compiled, a count was made, in both hospital and extra-mural locations, of all those with chronic disability following mental illness. Disability was defined as either continuous failure in

TABLE 9.7. *All mental illness: inceptions by marital state, sex, and age-group*

Marital state	*Average annual rate per 100,000 in 1961 census (numbers in parentheses), Salford, England (1959–63)* *Age-group*							
	15–19		*20–39*		*40–59*		*60+*	
Men								
Single	197	(53)	418	(140)	509	(62)	669	(28)
Married	0	(0)	248	(164)	202	(174)	294	(97)
Widowed	0	(0)	1,364	(3)	558	(16)	654	(57)
Divorced	0	(0)	370	(2)	315	(4)	1,538	(3)
Women								
Single	229	(58)	499	(101)	497	(68)	504	(55)
Married	346	(8)	400	(298)	255	(209)	281	(78)
Widowed	0	(0)	1,630	(11)	410	(45)	422	(164)
Divorced	0	(0)	978	(9)	333	(6)	1,356	(4)
Both sexes								
Single	213	(111)	448	(241)	503	(103)	550	(83)
Married	281	(8)	328	(462)	228	(383)	288	(175)
Widowed	0	(0)	1,564	(14)	441	(61)	464	(221)
Divorced	0	(0)	753	(11)	327	(10)	143	(7)

specified work and domestic role or as continuous residence in hospital.

Table 9.8 shows that among the chronically disabled the widowed were not over-represented when compared with inceptions, and with the 1961 census. Indeed, fewer were widowed than in either of the two age-adjusted comparison populations.

The contrast among the widowed between the high rate of inceptions and the low rate of chronic disability suggests that while bereavement may precipitate entry to psychiatric care, once the patients have entered care, widowhood does not lead to sustained disability. The result also counters the suggestion that an association between widowhood and mental illness arises because the widowed with chronic disability from mental illness are a residual group who do not remarry. A high rate of either recoveries or deaths among widows could have produced this

TABLE 9.8. *Patients with chronic disability following on mental illness: marital state and sex*

	Men			*Women*		
	*O**	*C**	*I**	*O*	*C*	*I*
Single	161	42	70	122	47	66
Married	68	199	159	108	180	153
Widowed	12	16	28	66	91	99
Divorced/separated	19	2	3	26	3	4
Total	260			322		

* Distribution by marital state and sex observed (O) in patients with chronic disability following on mental illness, compared with the distribution at the 1961 census (C), and at the inception of psychiatric care (I) (adjusted in each case to the age distribution of the population with chronic disability).

The expected numbers in the table have been rounded off, but in each case χ^2 was calculated from an expected number taken to two decimal points, and then itself rounded off.

result. An analysis of the deaths that occurred up to the end of 1965 among the inceptions of 1959–63 showed that it was not due to a high rate of deaths.

Transition to widowhood

To examine the effect of the transition into widowhood, the same information on the duration of widowhood was collected among three populations. Two of these were overlapping subsamples of the psychiatric register population. One subsample comprised all patients who entered psychiatric care in the calendar year 1962; they were subjects of a special survey of the complete patient population from Salford for that year. A second subsample comprised all patients entering care in the three-year period 1961–3, excluding those who were seen only in psychiatric out-patient clinics of general hospitals. Exactly comparable data on the general population were obtained from a special survey of a random sample of households in the community conducted in 1963. This sample comprised 1,981 adults.

Table 9.9 compares each of the two subsamples of patients with the household survey by duration of widowhood in three

TABLE 9.9. *Duration of widowhood: patients compared to survey population (percentage distribution by duration of widowhood in parentheses)*

Duration of widowhood	*All patients 1962*						*Patients 1961–3 excluding out-patients*						*Household survey*					
	Men		*Women*		*All**		*Men*		*Women*		*All†*		*Men*		*Women*		*All*	
1–12 months	2	(18)	4	(12)	6	(13)	5	(21)	11	(12)	16	(14)	2	(6)	7	(4)	9	(4)
1–5 years	3	(27)	13	(38)	16	(36)	8	(33)	20	(23)	28	(25)	10	(28)	28	(15)	38	(17)
5 or more years	6	(55)	17	(50)	23	(51)	11	(46)	57	(65)	68	(61)	24	(67)	148	(81)	172	(79)
Total no.	11	(100)	34	(100)	45	(100)	24	(100)	88	(100)	112	(100)	36	(101)	183	(100)	219	(100)

* Comparing distribution of duration of widowhood among *All patients* 1962 (all) and *Household survey* (all), $\chi^2 = 15{\cdot}43$, df = 2. The null hypothesis that duration of widowhood is not associated with inceptions of mental illness is rejected $P < 0{\cdot}005$.

† Comparing distribution of duration of widowhood among *Patients* 1961–3 (all) and *Household survey* (all), $\chi^2 = 15{\cdot}58$, df = 2. The null hypothesis that duration of widowhood is not associated with inceptions of mental illness is rejected $P < 0{\cdot}005$.

unequal time intervals. As compared with the household survey sample, both subsamples of patients show significant clustering in the earliest time interval after loss of the spouse, that is, in the first year. The patient populations were older than the subjects of the household survey, but this could not account for the clustering, for in each of the three samples the older age-groups had a smaller proportion than the younger of persons recently widowed. By inference, this time relation between the transition into widowhood and entry into psychiatric care provides strong support for a causal effect of the transition on entry to care.

Conclusion

We conclude from this analysis first that widowhood is associated with the inception of psychiatric care, an outcome which we believe reflects in fair degree the inception of severe mental illness. Second, widowhood is not associated with chronic disablement arising from mental illness. Third, among the widowed the inception of care for mental illness, and presumably the inception of mental illness itself, follow closely on the event of bereavement more often than expected by chance.

REFERENCES

1. Adelstein, A. M., Downham, D. Y., Stein, Z., and Susser, M. W. (1968). 'The epidemiology of mental illness in an English city: inceptions recognised by Salford psychiatric services', *Social Psychiat.* **3**, 47.
2. Kraus, A. S., and Lilienfeld, A. M. (1959). 'Some epidemiologic aspects of the high mortality rate in the young widowed group', *J. chron. Dis.* **10**, 207.
3. MacMahon, M., and Pugh, T. F. (1965). 'Suicide in the widowed', *Amer. J. Epidem.* **81**, 23.
4. Susser, M. W. (1968). *Community Psychiatry: Epidemiologic and Social Themes* (New York: Random House).

The use of psychiatric services by adolescents

A. S. HENDERSON

Department of Psychiatry, University of Tasmania, Australia
(formerly Member of the Scientific Staff, MRC Unit for
Research on the Epidemiology of Psychiatric Illness, Edinburgh)

There is a paucity of information on how much psychiatric morbidity the teenage years carry. For children, there are the studies of Pasamanick (15), of Lapouse and Monk (14), and of Rutter and Graham (17). For old people there are studies such as the Newcastle upon Tyne survey by Kay, Beamish, and Roth (11) and Gruenberg's (7) work in Syracuse. For adults in general, we are all familiar here with many epidemiological studies at the level of the hospital, the family doctor, and the community itself. To my knowledge there has as yet been no survey of morbidity in adolescents either at the field survey or the family doctor level, yet the teenage years are generally accepted to constitute a period of special stress.

McCulloch, Philip, and I carried out what we should like to consider as the first stage of an *epidemiological* inquiry into adolescent psychiatric illness. We studied the use made of the psychiatric services in Edinburgh by persons aged 14–19 years inclusive, during a one-year period in 1964–5. We collected data on all persons who met the following criteria: (1) they had had a disturbance in mental health resulting in at least one consultation with a psychiatrist in the study year; (2) they were aged 14–19 years inclusive at the time of that first consultation; (3) they were resident in the City of Edinburgh on the night prior to the consultation; (4) they were examined in the twelve-month period July 1964 to June 1965, having had no consultations in the previous five years; (5) they had a measured or estimated IQ of 70 or over.

On the basis of the 1961 census, there were 41,000 teenagers at risk in our chosen age-group in the City of Edinburgh. In the survey year 230 teenagers were examined by a psychiatrist. This represents an annual referral rate of 5·6 per thousand at risk. Approximately half were seen only once (42 per cent of boys and 56 per cent of girls). One-half had been referred by their family doctor (58 per cent of boys and 45 per cent of girls). One-third had carried out an act of self-poisoning or self-injury immediately before reaching consultation (30 per cent of boys and 37 per cent of girls).

This latter group revealed some interesting characteristics on closer examination. My colleagues calculated for each of the 23 wards of the city the rates for teenagers currently in psychiatric treatment, i.e., the ward rates for new referrals plus those already in treatment. When these rates were ranked for persons reaching help after self-poisoning or self-injury, and after orthodox referral, there were highly significant differences. There was an inverse relationship between rates for self-poisoning or self-injury and rates for *direct* referral to a psychiatrist. Furthermore, those adolescents who were referred by their general practitioner came from the more affluent, socially integrated areas of the city, while those who poisoned or injured themselves came from areas conspicuous in those social variables indicating disorganization, such as overcrowding, rent arrears, and delinquency (Kendall's Tau $= 0{\cdot}2806$, $p < 0{\cdot}05$). The self-poisoning group was considered by psychiatrists to have had predominantly acute reactions to stress: the commonest diagnoses used were 'nil psychiatric' and 'adolescent crisis'. In contrast, the group coming to consultation by orthodox means (usually referral by the family doctor) were considered to have disorders of longer duration, such as neurosis, behaviour disorders, or abnormal personality. One is aware of the biases likely to operate in the use of such diagnoses by a psychiatrist, particularly those biases operating when a patient from a deprived subculture, very different from the psychiatrist's own, is being examined after a self-poisoning act. For this reason, Philip's and McCulloch's (1966) psychometric study of attempted suicide should be repeated in this age-group where *transient* emotional disturbances are believed to be specially common. This would provide an objective measure of the psychiatric state of adolescents immediately after recovery from a self-poisoning act.

One interpretation of the findings in this modest and introductory investigation of adolescent illness is as follows. There are reasons for believing that, at the field level, rates for formal psychiatric diagnoses among adolescents—neuroses, personality or behaviour disorders, or adolescent crises—really do vary substantially from one subculture or social class to another. Yet the preludes to acts of self-poisoning, such as rows with parents or boy-friends, are likely to occur uniformly in all strata of society. Such stresses for the adolescent are probably dealt with by different methods across the classes. For one subculture, usually in the disorganized sector, there is an impulsive response. In another subculture there may be more rational communication and therefore dissipation of the adolescents' turbulent emotions. The teenager and his family can handle such transient feelings. When a teenager shows signs of distress or unacceptable behaviour, more use seems to be made of the family doctor. This does not apparently occur nearly so readily in the socially disorganized subculture. In the U.K., hesitation in using a doctor for this purpose does not arise through financial restrictions.

Students of mental health in adolescents, such as Erikson (3, 4, 5), Anna Freud (6), and Blos (2) have emphasized that the teenage years are particularly stressful. I believe it is an area ripe for careful epidemiological inquiry. We asked ourselves if there was any evidence in our Edinburgh study of an excess of specialist-treated illness in teenagers compared with *adults*. We compared our annual referral rate of 5·6 per thousand with the best obtainable data—that for adults elsewhere in the U.K. (Table 9.10).

The conclusion to be made is that turmoil there may be but there is no excess of formal mental illness reaching the psychiatric services. Conceivably, there are many teenagers who experience symptoms but who do not report these—much as happens in the elderly. We do not yet know the circumstances in which an adolescent or his parents seek help from a general practitioner. Nor do we know the extent to which a general practitioner recognizes psychiatric illness in this age-group. We do know, from the general practice studies by Kessel (12) and by Shepherd *et al.* (18) that once psychiatric illness is recognized, referral to a psychiatrist is more likely for young patients than for their elders.

TABLE 9.10

Authors	Community	Annual referral rate per thousand at risk*
Adults		
Hare and Shaw (9)	Croydon	5·8 (new area) 4·7 (old area)
Kessel (13)	Plymouth	7·8 (1961 figure) 8·4 (1964 figure)
Grad and Sainsbury (8)	Chichester	6·8
	Salisbury	5·3
Baldwin (1)	North-east Scotland	8·7
Adolescents		
Henderson, McCulloch, and Philip (10)	Edinburgh	5·6

* Some authors' figures include persons down to the age of 15 or 16 years in the over-all adult figure.

In the epidemiological paradise of Tasmania, I am attempting to evolve a research strategy to extend this work. The most rewarding approach would seem to be to determine the prevalence of formal psychiatric symptoms among adolescents at the field survey level and to determine some of the social and personality characteristics of the high- and low-risk groups.

REFERENCES

1. Baldwin, J. A. (1967). Personal communication.
2. Blos, P. (1962). *On Adolescence* (New York: Free Press of Glencoe).
3. Erikson, E. H. (1956). 'The problem of ego identity', *J. Amer. psychoanal. Ass.* **4**, 56–121.
4. —— (1961). 'The roots of virtue', in *The Humanist Frame*, ed. Huxley, J. (London: Allen and Unwin).
5. —— (1962). 'Youth, fidelity and diversity', *Daedalus*, **91**, 5.
6. Freud, A. (1958). 'Adolescence', in *The Psychoanalytic Study of the Child*, vol. 13, ed. Eissler (New York: International Universities Press).
7. Gruenberg, E. M. (1961). *A Mental Health Survey of Older People* (New York: Utica).
8. Grad, J., and Sainsbury, P. (1965). 'Evaluating the community psychiatric services in Chichester'. *Milbank mem. Fd Quart.* **44** no. 1.
9. Hare, E. H., and Shaw, G. K. (1965). *Mental Health on a New Housing Estate: A Comparative Study of the Health in Two Districts of Croydon* (London: Oxford University Press).

10. HENDERSON, A. S., McCULLOCH, J. W., and PHILIP, A. E. (1967). 'Survey of mental illness in adolescence', *Brit. med. J.* **1**, 83–4.

11. KAY, D. W. K., BEAMISH, P., and ROTH, M. (1964). 'Old age mental disorders in Newcastle-upon-Tyne', *Brit. J. Psychiat.* **110**, 146–58.

12. KESSEL, W. I. N. (1960). 'Psychiatric morbidity in a London general practice', *Brit. J. prev. soc. Med.* **14**, 16–22.

13. —— (1965). 'The evaluation of the Plymouth Nuffield Clinic', *Milbank mem. Fd Quart.* **44**, no. 1.

14. LAPOUSE, R., and MONK, M. A. (1958). 'An epidemiologic study of behaviour characteristics in children', *Amer. J. publ. Hlth,* **48**, 1134–44.

15. PASAMANICK, B. (1956). 'The epidemiology of behaviour disorders in children', *Neurology and Psychiatry in Childhood,* **34**, 397.

16. PHILIP, A. E., and McCULLOCH, J. W. (1966). 'Use of social indices in psychiatric epidemiology', *Brit. J. prev. soc. Med.* **20**, 122.

17. RUTTER, M., and GRAHAM, P. (1966). 'Two surveys of children: psychiatric disorder in 10- and 11-year old children', *Proc. roy. Soc. Med.* **59**, 382–7.

18. SHEPHERD, M., OPPENHEIM, A. H., and MITCHELL, S. (1966). 'Two surveys of children: the definition and outcome of deviant behaviour in childhood', ibid. **59**, 379–82.

Spatial distribution of mental disorders in Mannheim, 1965

HEINZ HÄFNER and HELGA REIMANN
Department of Social Psychiatry, University of Heidelberg, West Germany

The object of this investigation was to discover whether the associations found by Faris and Dunham (3) and others, between the frequency of mental disorders and ecological factors such as poverty or social disorganization, were also characteristic of the German city of Mannheim. We also hoped to obtain data which would help in planning and evaluating a new community mental health centre.

Material and method

We counted all inhabitants of Mannheim who in 1965 first consulted certain specified services because they were suffering from a mental disorder or mental retardation (this is our definition of 'incidence'). The agencies included were as follows:

(*a*) The 307 state, university, and private mental hospitals, general hospitals, and private sanatoria in Germany and Switzerland.

(*b*) All twelve psychiatric practitioners in Mannheim.

(*c*) The Public Health Office of Mannheim.

(*d*) All forms of public, church, and private welfare provision, such as old people's and children's homes, educational facilities, family, and marriage counselling services, associations of alcoholics, etc.

Case identification was made from the records kept by these agencies. The diagnoses of individuals attending non-professional services were assessed on the basis of the symptoms reported, except in cases where validity was not in doubt, such as mental retardation, alcoholism, and senile brain diseases. We accepted considerable diagnostic uncertainty because of the

TABLE 9.11. *Comparison of the annual incidence rates of mental disorders found in four different studies*

Authors	*Census year*	*Community*	*Population*	*Rate per thousand*
Hollingshead and Redlich (5)	1950	New Haven (U.S.)	236,940	2·08
Hagnell (4)	1957	Lundby (Sweden)	2,421	7·48
Adelstein, Downham, Stein, and Susser (1)	1961	Salford (England)	117,374	3·3
Reimann and Häfner	1965	Mannheim (Germany)	328,106	10·74*

* The diagnoses and the methods used in the Mannheim project do not exactly correspond to those applied in the three other studies. The Mannheim study, as well as the studies of New Haven and of Lundby, have included patients of all ages, thus covering the first utilization for mental disorders of children under 15, whereas the Salford study only comprises the first utilizations of the population aged 15 and over. The Mannheim study also includes first consultations for mental retardation.

advantage of including a large number of services. The reliability of the results can be partly assessed by comparing the first consultation rate of psychiatric services with that of welfare provisions and by the separate calculation of associations with the basic ecological variables (Table 9.11).

The annual incidence rate in 1965 for all psychiatric disorders in Mannheim is relatively high at 10·74 per thousand. Not only the studies noted in the table, but also other approximately comparable studies, for example those by Norris in London (11), Klee *et al.* in Baltimore (7), and Walsh in Dublin (16), showed lower first admission rates. This can be explained by the fact that, contrary to most of the comparable studies, the rate of 1·53 per thousand for mental retardation is included in our results, and that contacts with many non-hospital social services were added.

Mannheim is an industrial and university town with about 330,000 inhabitants, situated 12 miles from Heidelberg where the river Neckar flows into the Rhine. It was founded in 1606 by the Elector Frederic IV and became the fortified residential town of the Palatinate in place of Heidelberg until 1770. The present city is divided into exact rectangles. It is dominated by the large castle built in 1720 which nowadays is the seat of the university. The Elector Charles Philippe decided that the highest dignitaries at court had to live in the immediate vicinity of the castle and the lower ranks at an increasing distance away from it. It is regrettable that very little remains of this spatial distribution, so valuable for the epidemiologist. Fig. 9.2 gives a topographic survey of some crude spatial characteristics of the city.[1]

Comparable to the Chicago study by Park and Burgess (12), whose ecological data were used by Faris and Dunham, the city contains centres of administration, traffic, and business.

The housing areas in the centre (01–04) consist to a large extent of lower-middle standard apartment houses. In the north of the city is situated the large old harbour surrounded by an industrial area. This area is adjoined by housing areas (05) with mostly old low-standard dwellings and signs of social disorganization, especially in the corner between the Neckar and the Rhine.

The second big harbour area, which was built much later than the first, about the turn of the century, is situated in the south of the city and is surrounded by industrial estates. Its environment is dominated by workmen's settlements, small private houses, and apartment houses of lower-middle standard. In the south-east, housing areas of upper-middle standard, and upper standard, such as villas and apartment houses, adjoin the city centre. Towards the north-east are older apartment houses, plain government-subsidized dwellings, and small houses of lower-middle standard. On the outskirts of this area, workmen's settlements of lower-middle standard have arisen around several incorporated villages which up to now lead a relatively active community life.

Tract 13 (Speckweg-West) and tract 29 (Hochstätt) occupy a special position. Tract 13 is situated in the neighbourhood of an industrial area and mainly consists of slums of low housing standard (partially barracks) and with distinct social disorganization. Tract 29, in the south-east, came into existence through a government-subsidized housing programme only about six years ago. Certain families from other districts of Mannheim

---

[1] Our rating of urban districts is based on information obtained from the Building Surveyor's Office and the Social Welfare Office of the City of Mannheim.

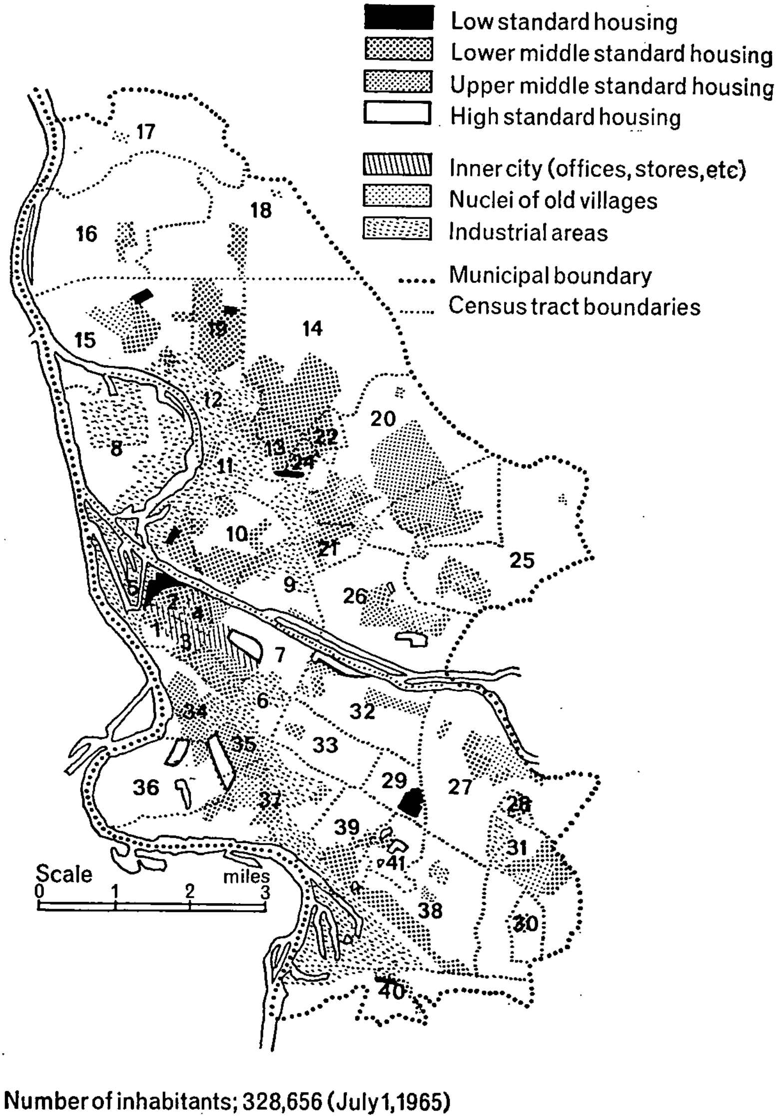


FIG. 9.2. Crude standards of housing in Mannheim

were assigned these new cheap lower-middle standard dwellings because they lived in lodgings which were ready for demolition or were not able to pay higher rents or were problem families.

First utilization rates of mental disorders in Mannheim, 1965

Forty-one census tracts were combined, on the basis of spatial contiguity and approximately similar incidence rates, into 20 urban districts. Within these groupings, the annual incidence rate of all mental disorders varied between 5·82 and 17·86 per thousand, with a mean of 10·74 (Table 9.12).

The remaining Figs. 9.3–9.6 show the geographical distribution of mental disorders, divided into five groups according to size of incidence rate. Fig. 9.3 shows that the highest total incidence rates are to be found in the city centre, in the slum area 05 (Jungbusch) adjacent in the north-east, and in the two smaller slums 13 (Speckweg-West) and 29 (Hochstätt) in the south-east and north-east of the city. A slight increase in these rates is to be found in a southern subcentre (38–41, Rheinau), in the more recently built industrial area around the second harbour.

The rates are slightly increased in the industrial and housing area 34 (Lindenhof) which is adjacent to the city centre in the south with lower, middle, and higher standard housing. It is striking that this is also the case with the residential area 07 (Oststadt) which consists of middle and higher standard housing bordering the city in the south-east.

The irregular belt of mainly middle standard residential areas which surrounds the city presents incidence rates slightly below average. The lowest rates are to be found in the middle and higher standard area 32–3 (Neuostheim) as well as in the village communities on the outskirts of the city.

In order to consider separate diagnostic groups, the census tracts were combined still further into twelve city districts according to the same principles of geographical contiguity and similarity of rates used earlier.

At 53·6 per 100,000 the crude incidence rate of schizophrenia in Mannheim is more than twice as high as the mean of 21·8 per 100,000 calculated in 1965 by Dunham from different studies,

TABLE 9.12. *First utilization rates of all mental disorders in Mannheim, 1965, divided into 20 urban districts*

Census tracts	*Urban districts*	*Ill persons per thousand inhabitants*
01		
02		
04	1. Innenstadt	13·64
03	2. Östl. Oberstadt	17·86
05	3. Jungbusch-Mühlau	16·45
06	4. Schwetzingerstadt	10·35
07	5. Oststadt	12·86
08		
09		
10	6. Neckarstadt	10·36
11		
12	7. Waldhof	8·66
13	8. Speckweg-West	15·60
14		
15		
16		
17		
18	9. Sandhofen	7·62
19	10. Schönau	6·64
20		
21		
22		
23		
24	11. Käfertal-Gartenst	9·65
25	12. Wallstadt	5·82
26	13. Feudenheim	9·70
27		
28	14. Seckenheim	6·68
29	15. Hochstätt	13·27
30		
31	16. Friedrichsfeld	7·94
32		
33	17. Neuostheim-Neuhermsh	7·59
34	18. Lindenhof	10·91
35		
36		
37	19. Neckarau	8·29
38		
39		
40		
41	20. Rheinau	11·38
Mannheim total		10·74*

Differences significant ($p < 0{\cdot}05$) by χ^2 test.

* First consultations for mental retardation are included.

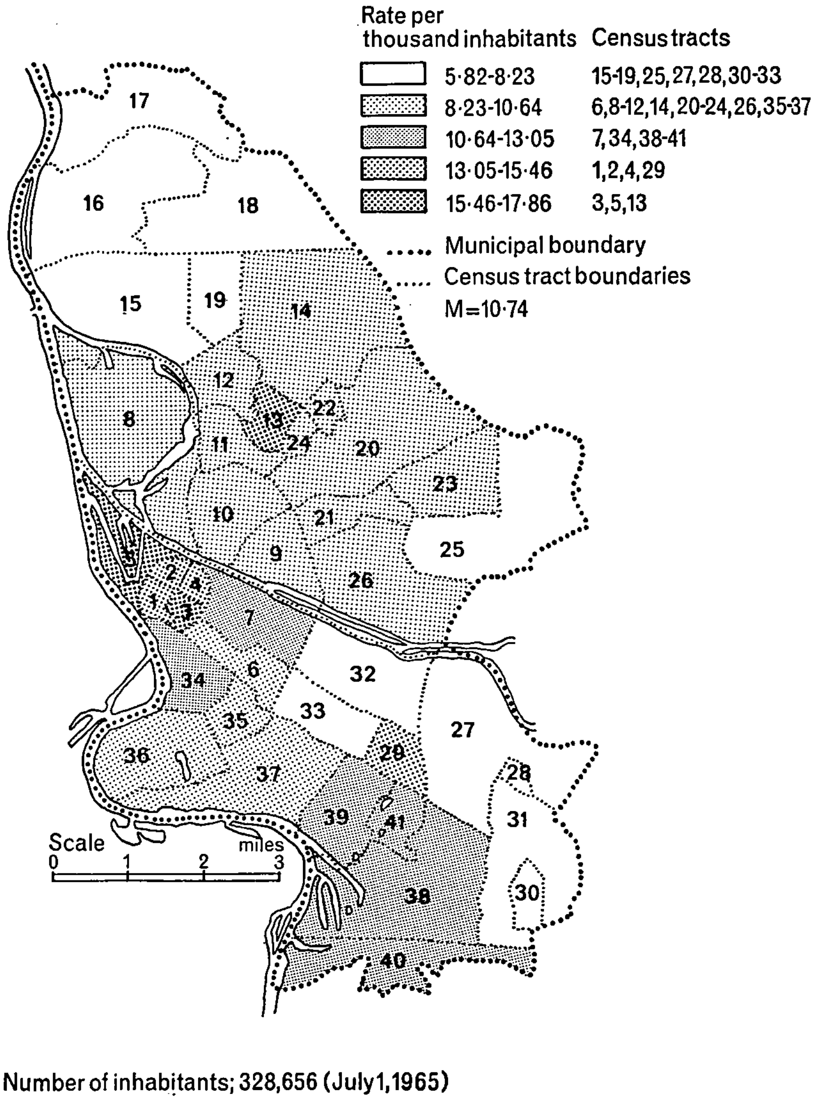


FIG. 9.3. First utilization rates of all cases of mental disorder

and two to three times as high as the rates of 23·8 or 15·8 respectively, quoted by Kramer for the U.S.A. and England and Wales in 1969. It corresponds, however, to the rate of 52 per 100,000 given by Walsh for Dublin in 1969.

Fig. 9.4 shows that the rates are highest in the city and in other lower standard and socially disorganized areas. The lowest rates are to be found on the outskirts, particularly in the village areas with strong community activities.

The crude incidence rate for depressive psychoses in Mannheim, 81 per 100,000 (Fig. 9.5) is higher than the rate of 27·1 given by Kramer in 1969 for England and Wales which, according to his statement, is already about 800 per cent higher than the comparable rate of 3·0 for the US. The topographic distribution shows, as expected, no clear association with ecological variables.

The largest diagnostic group consists of neuroses, personality and psychosomatic disorders, and emotional crises, behaviour disturbances and psychopathic disorders of children (Fig. 9.6).

TABLE 9.13. *First utilization rates of mentally retarded children (age 0–14)*

Census tracts	*Population 1965 (age 0–14)*	*Retarded children (age 0–14) number of cases*	*Rate per thousand age-standardized*
01–04	4,123	27	6·55
05	1,218	11	9·03
06–07	4,496	23	5·12
08–10	9,274	41	4·42
11–13	4,868	71	14·59
15–18	3,178	8	2·52
14, 19–24	11,317	123	10·87
25–6	3,923	12	3·06
27–33	4,835	20	4·14
34	2,665	11	4·13
35–7	7,857	16	2.04
38–41	4,856	59	12·15
Unknown residency		4	
Mannheim total	59,945	421	7·02

$\chi^2 = 149{\cdot}58$. df = 11. $p < 0{\cdot}001$ highly significant distribution.

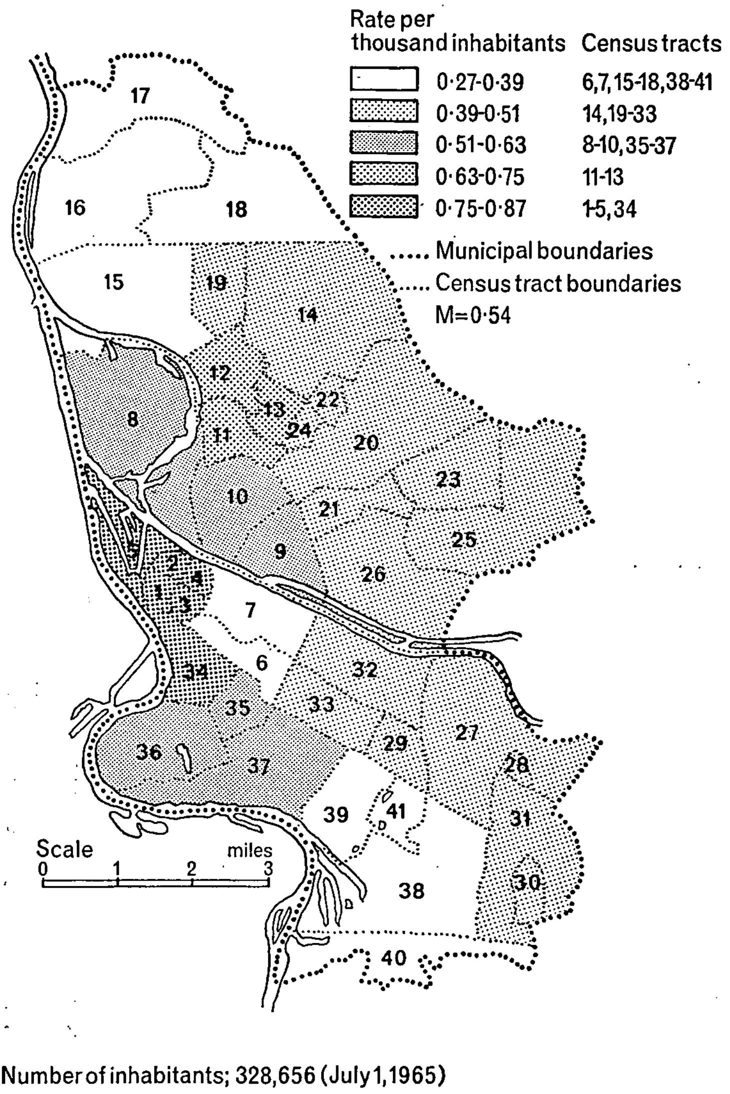


FIG. 9.4. First utilization rates of schizophrenia

With a rate of 3·65 per thousand this group amounts to seven times the rate for schizophrenia and equals in size the prevalence rates found in medical general practice (6, 9, 14, 15). Its topographic distribution coincides to a large extent with that found for all mental disorders.

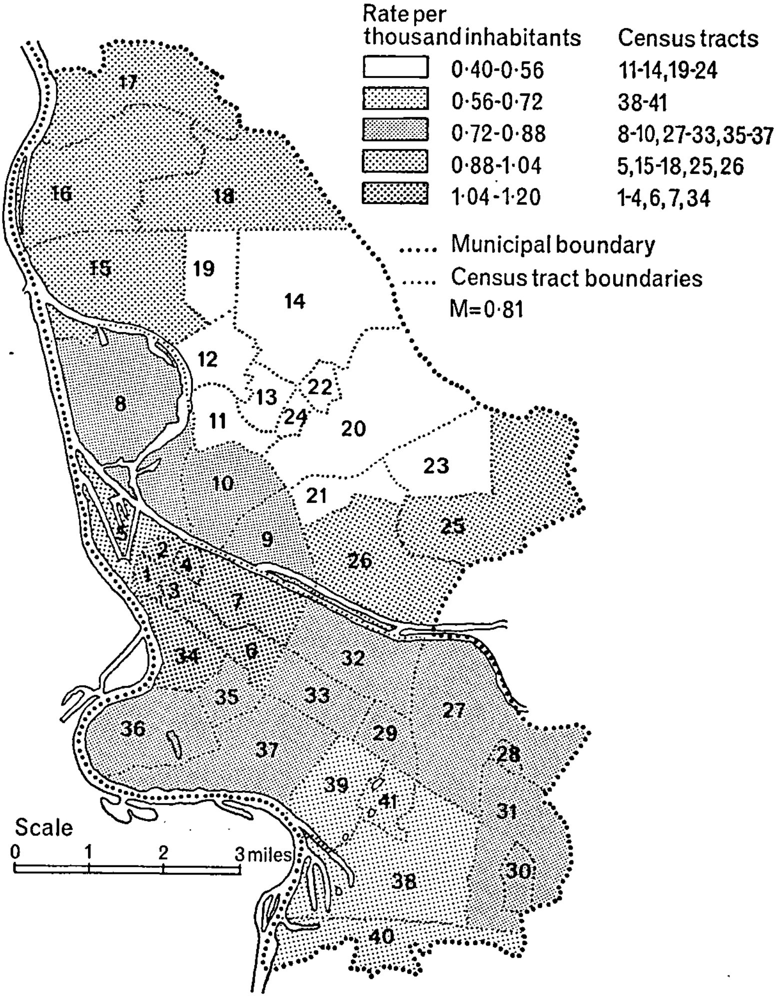
Eighty-seven per cent of the registered cases of mental retardation were aged under 15. Here the number of individuals in contact with services is somewhat more closely connected to the true incidence. Compulsory education forces the parents to send each child to school. Children who are not mature enough to go to school or are not able to keep up with the elementary school due to a lack of intelligence are tested and advised by the psychological service of the school board or by the Public Mental Health Office. Thus, one can expect that the loss by non-consultation of services is relatively small, at least in the past ten years.

Table 9.13 shows the age-standardized incidence rates of

TABLE 9.14. *First utilization rates of arteriosclerotic and senile brain diseases in inhabitants of Mannheim aged 50 and over*

Census tracts	*Population 1965 aged 50 and over*	*Arteriosclerotic and senile brain diseases number of cases*	*Rate per thousand age-standardized*
01–04	10,008	63	6·2950
05	2,343	23	9·8165
06–07	10,577	63	5·9563
08–10	21,426	111	5·1806
11–13	5,668	17	2·9993
15–18	5,170	19	3·6750
14, 19–24	14,322	40	2·7929
25–6	15,525	34	2·1900
27–33	7,226	23	3·1830
34	6,527	19	2·9110
35–7	11,385	46	4·0404
38–41	6,030	27	4·4776
Unknown residency		9	
Mannheim total	116,207	494	4·2510

$\chi^2 = 76$. df = 11. $p < 0{\cdot}001$ highly significant distribution.



Number of inhabitants; 328,656 (July 1, 1965)

FIG. 9.5. First utilization rates of depressive and manic psychoses

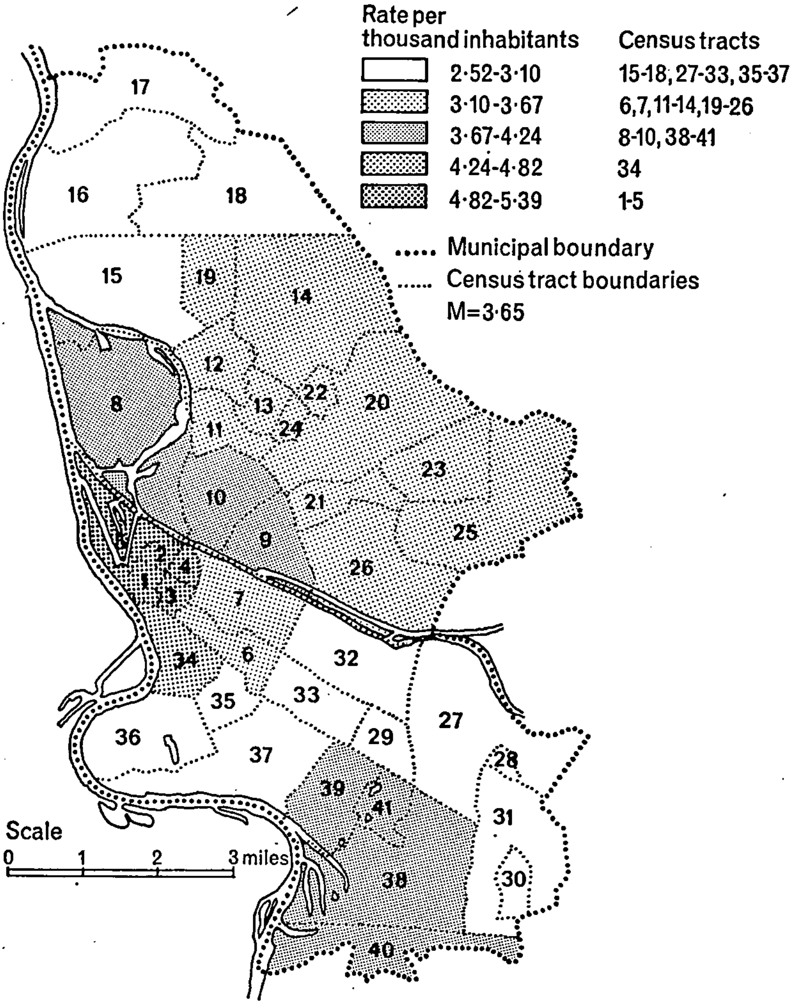
twelve city districts, They vary between 2·04 and 14·59 per thousand, that is, by the factor 7 approximately. The topographic distribution is surprising. Only four districts show high or very high rates (higher than 9 per thousand): the slum area 05 (Jungbusch), the relatively new suburban area for factory and office workers 14, 19–24 (Schönau, Käfertal, Gartenstadt), the southern workmen area 38–41 (Rheinau), and the slum area 11–13 (Speckweg-West). The high rate in the mixed district 14, 19–24 (Schönau, Käfertal, Gartenstadt) is partially due to a smaller slum and low standard area (in 22 and 24) adjacent to district 13 (Speckweg-West) and to a smaller zone of substandard housing in 19 (Schönau).

The age-standardized utilization rates for arteriosclerotic and senile brain diseases (Table 9.14) show smaller topographical differences. The highest rates are to be found in the slum district 5 (Jungbusch) near the city and also, though not quite so high, in the city itself. The rates in the upper and middle standard region 6–7 (Oststadt) are fairly high, whereas the slum region 11–13 (Speckweg-West, etc.) presents strikingly low rates. The first utilization rate depends on numerous social variables which have not yet been evaluated, such as living alone or in a family setting, attitudes, housing, etc.

Discussion

These results are preliminary, and it is too soon to offer interpretations. There is a tendency for higher rates of first utilization for schizophrenia and mental retardation to be associated with unfavourable ecological conditions, such as low standard housing and social disorganization, while psychotic depressions show no such relationship. These facts are to be expected on the basis of earlier studies.

The similarity of the topographic distribution to that found by Faris and Dunham in the Chicago study is interesting. Through industrialization and migration, comparable distribution processes seem to have taken place in Mannheim, in spite of its different history. These processes have led to the formation of varying ecological sub-regions. Further investigation of the distribution of mental disorders, and of the reasons for high rates in certain areas, is under way.



Number of inhabitants; 328,656 (July 1, 1965)

FIG. 9.6. First utilization rates of neurotic disorders

REFERENCES

1. ADELSTEIN, A. M., DOWNHAM, D. Y., STEIN, Z., and SUSSER, M. W. (1968). 'The epidemiology of mental illness in an English city', *Social Psychiatry,* **3**, 47.
2. DUNHAM, H. W. (1965). *Community and Schizophrenia: An Epidemiological Analysis* (Detroit: Wayne State University Press).
3. FARIS, R. E. L., and DUNHAM, H. W. (1939). *Mental Disorders in Urban Areas* (Chicago: University of Chicago Press.)
4. HAGNELL, O. (1966). *A Prospective Study of the Incidence of Mental Disorder* (Stockholm: Svenska Bokförlaget Norstedts-Bonniers).
5. HOLLINGSHEAD, A. B., and REDLICH, F. C. (1958). *Social Class and Mental Illness* (New York: Wiley).
6. KESSEL, W. I. N. (1960). 'Psychiatry morbidity in a London general practice', *Brit. J. Prev. soc. Med.* **16**, 14.
7. KLEE, G. D., SPIRO, E., BAHN, A. K., and GORWITZ, K. (1967). 'An ecological analysis of diagnosed mental illness in Baltimore', in *Psychiatric Epidemiology and Mental Health Planning,* eds. Monroe, R. P., Klee, G. D., and Brody, E. B. (Psychiatric Research Reports of the American Psychiatric Association No. 22, Washington).
8. KRAMER, M. (1969). 'Cross-national study of diagnosis of mental disorders: origin of the problem', *Amer. J. Psychiat.* **125**, Supp. 1.
9. LOCKE, B. Z., FINUNANCE, D. L., and HASSLER, F. (1967). 'Emotionally disturbed patients under care of private non-psychiatric physicians', ibid.
10. MORRIS, J. N. (1969). *Uses of Epidemiology* (revised edition) (Edinburgh: E. and S. Livingstone).
11. NORRIS, V. (1959). *Mental Illness in London.* Maudsley Monograph No. 6, Institute of Psychiatry (London: Chapman and Hall).
12. PARK, R. E., and BURGESS, E. W. (1925). *The City* (Chicago: University of Chicago Press).
13. ROVITZ, L., and LEVY, L. (1968). 'Ecological analysis of treated mental disorders in Chicago', *Arch. gen. Psychiat.* **19**, 571.
14. SHEPHERD, M., COOPER, B., BROWN, A. C., and KALTON, G. W. (1964). 'Minor mental illness in London: some aspects of a general practice survey', *Brit. med. J.* **2**, 1359.
15. STROTZKA, H. (1969). *Kleinburg: Eine Sozialpsychiatrische Feldstudie* (Wien: Oesterreichischer Bundesverlag).
16. WALSH, D. (1969). 'Mental illness in Dublin: first admissions', *Brit. J. Psychiat.* **115**, 449.

Subcultural aspects of attempted suicide

NORMAN KREITMAN
MRC Unit for Epidemiological Studies in Psychiatry, University Department of Psychiatry, Edinburgh

The study to be described arose out of two considerations. First, it appeared clinically that many patients admitted to hospital for so-called 'attempted suicide' were members of social groups in which suicidal behaviour of a similar kind was unduly common, even when allowance was made for social class, overcrowding, unemployment, and other variables known to be associated with high rates of attempted suicide (4). Secondly, it seemed that the 'cry for help' aspect of attempted suicide, which has been ably documented by Stengel (6) and by Farberow and Shneidman (1), could only be understood as a social rather than as a purely individual phenomenon, and if a word or deed is to serve as a communicational symbol, i.e., a language, it must be employed relatively more often by all those within the communicational system than by those outside it. The concept of attempted suicide as a language would be strengthened by a demonstration that such acts occurred with unexpected frequency in the patient's social orbit.

These considerations led to the hypothesis that among the social entourage of patients admitted to hospital for suicidal attempt there would be a higher prevalence of attempted suicide than would be expected from a comparable group of individuals in the general population.

Method

The study was carried out at the Regional Poisoning Treatment Centre (RPTC), Royal Infirmary, Edinburgh, which is known to receive over 90 per cent of all hospital-treated cases in the city (2). Essentially, the procedure consisted of questioning a series of patients (probands) admitted to hospital following a suicidal act, in order to identify the individuals in their social set (contacts) and to obtain data on the contacts' age, sex, current and previous addresses. Attention was confined to specified categories of kinship which included spouse, parents, siblings, children, the spouses of sibs, and of children, and two intimate friends. The data were checked by similarly questioning another member of the family of 10 per cent of the probands and were found to be of acceptable reliability. Contacts who were aged 15 or over and who were resident in the city of Edinburgh for at least part of the four-year period, 1963–6, comprised the contact population ($n = 578$).

The next step was to calculate for *each* contact his expectation of having been admitted to hospital with a suicidal attempt at least once over the four-year study period, 1963–6. This interval was used as it corresponds to the time for which detailed epidemiological data for Edinburgh have been available. In determining the expectation of admission, consideration was given to sex, age, duration of risk, i.e., proportion of the four years spent at an identified address or addresses within the city, and an 'area factor'. These area factors had been computed as standardized morbidity ratios for each of the city wards and represented the combined effect of a wide range of social variables known to be correlated with rates for attempted suicide. They are, of course, essentially ecological indexes, but their usefulness and validity for present purposes have been explored by other studies (4, 5).

Subsequently it was possible to determine by simple summation the total number of admissions to be expected among the contact population, or any subgroup thereof, over the study period.

This expected number was then compared with the number of individuals in the contact population identified from the hospital registers as attempted suicides for the four-year period.

Full details of the method, which includes a computer simulation exercise to estimate the magnitude of error for certain estimates, has been published elsewhere (3).

Results

Seventeen members of the contact population were admitted to the RPTC during the study period, which is significantly more than the expected number of 4·28 ($p < 0{\cdot}001$).

Younger patients, below 35 years of age, had a positive contact significantly more often than older patients ($p < 0{\cdot}02$), the age effect being particularly marked for women ($p < 0{\cdot}05$) although sex considered alone did not have an unequivocal effect (see Table 9.15). Patients who had attempted suicide by drugs rather than self-injury also had a relative excess of positive contacts.

TABLE 9.15. *Proband characteristics: age and sex and distribution of positive contacts*

Proband characteristics	*Number of contacts at risk*	*Number observed* (O)	*Number expected* (E)	*Proportion expected* (E^1)
Sex: Male	269	8	2·11	8·48
Female	309	9	2·12	8·52
Total	578	17	4·23	17·00
	$\chi^2 = 0{\cdot}054$			
	df = 1			
	$p < 0{\cdot}80$			
Age: −34	299	14	2·27	8·93
35+	279	3	1·96	8·07
Total	578	17	4·23	17·00
	$\chi^2 = 6{\cdot}063$			
	df = 1			
	$p < 0{\cdot}02$			
Age and sex:				
Male −34	135	5	1·17	4·70
35+	134	3	0·94	3·78
Female −34	164	9	1·10	4·42
35+	145	0	1·02	4·10
Total	578	17	4·23	17·00
	$\chi^2 = 9{\cdot}026$			
	df = 3			
	$p < 0{\cdot}05$			

There was no particular category of contact which yielded a significant excess of 'positives' and, in particular, blood kin, i.e., parents, sibs, and children, did not contain proportionally more positives than kin genetically unrelated to the proband, such as in-laws.

Among the seventeen patient-contact pairs there was a tendency for the interval between the admissions of the two individuals to be shorter than might be expected on a chance basis, though the small numbers preclude any firm conclusion. Inspection of the clinical records revealed no cases in which both individuals of a patient-contact pair had carried out their self-aggressive acts in response to a common environmental stress.

Discussion

Methodologically the central problem in this study had been the determination of the expectation of admission for each of the individuals in the contact population. The principles are simple, and the age, sex, and duration-at-risk components presented no problems, but the use of area factors is more debatable. It had been established from earlier studies that the patients are typical of the areas from which they are admitted in that the characteristics of social disorganization of their areas are at least equally prevalent in their own environments, in so far as such assessments can be made. The patients may, however, show such characteristics to an even greater degree than their neighbours, in which case the area factors will tend to underestimate the stresses falling on them and presumably on their families. To this degree the expected number of admissions in the contact population will have been underestimated. Against this may be offset the consideration that area factors may well reflect cultural features of the local communities, such as its social structure, which cannot readily be defined in terms of variables applicable to individuals. However, enough uncertainty prevails to necessitate some caution in accepting the findings of this study.

The principal finding was the fourfold excess of the observed over the expected number of positive contacts. This is certainly compatible with the hypothesis that a subcultural milieu is important. Such a milieu appears to be particularly important for young women and this is reminiscent of what is known about *folie-à-deux* and epidemic hysteria. The findings are also compatible with the view that within the subculture a suicidal attempt may act as a communication. It could be argued that only one form of communication has been employed here, namely, that of similar or symmetrical responses. Such a pattern seems likely to be the exception rather than the rule in any human communication system and other types of behaviour are doubtless more typical antecedents or consequences of suicidal attempts. Their delineation and study might contribute not only to a further understanding of attempted suicide but of pathological forms of communication in general.

REFERENCES

1. Farberow, N., and Shneidman, E. (1961). *The Cry for Help* (London: McGraw-Hill).
2. Kessel, N., McCulloch, J. W., Hendry, J., Leslie, D., Wallace, I., and Webster, R. (1964). 'Hospital management of attempted suicide in Edinburgh', *Scot. med. J.* **9**, 333.
3. Kreitman, N., Smith, P., and Tan, E. (1969). 'Attempted suicide in social networks', *Brit. J. prev. soc. Med.* **23**, 116–23.
4. McCulloch, J. W., and Philip, A. E. (1967). 'Social variables in attempted suicide', *Acta psychiat. scand.* **43**, 341.
5. Philip, A. E., and McCulloch, J. W. (1966). 'Use of social indices in psychiatric epidemiology', *Brit. J. prev. soc. Med.* **20**, 122.
6. Stengel, E., and Cook, N. (1958). *Attempted Suicide* (London: Oxford University Press).

The effect that coroners have on the suicide rate and the open verdict rate

B. M. BARRACLOUGH
MRC Clinical Psychiatry Research Unit, Chichester, Sussex

Suicide statistics in England and Wales are compiled from death certificates completed by coroners following an inquest. The decisions represented by the certificates are neither subjected to scrutiny nor, except in rare instances, revised. This paper describes an inquiry designed to examine firstly, whether —as is frequently alleged—the personal idiosyncrasies of coroners influence these decisions and hence the suicide rate and secondly, whether there is evidence of a common standard among coroners as to what a suicide is and what an open verdict is.

The county boroughs of England and Wales are urban areas each having its own coroner, its own census return, and its own health statistics. The coroners are usually lawyers and are appointed by the county council for life with a salary. As judicial officers they are responsible to the Crown and not to the appointing authority, and therefore local affairs need not affect their deliberations.

Fig. 9.7 shows the frequency distribution of the three-year average crude suicide rate (per million) for 79 county boroughs for 1950–2. It is a skewed Gaussian curve. There is a similar curve for 1960–2 and the correlation coefficient for the rates between the periods is 0·45.

Between 1953 and 1959, 19 coroners were replaced. We made the hypothesis that if the personal idiosyncrasies of the coroner had an effect on the decision as to what a suicide was, then the correlation coefficient between the two periods would be significantly lower in those boroughs where the coroner had changed than in those where he had not. The coefficients were respectively 0·49 and 0·45, thus refuting the hypothesis that the personal idiosyncrasies of the coroners affect the suicide rate (Table 9.16).

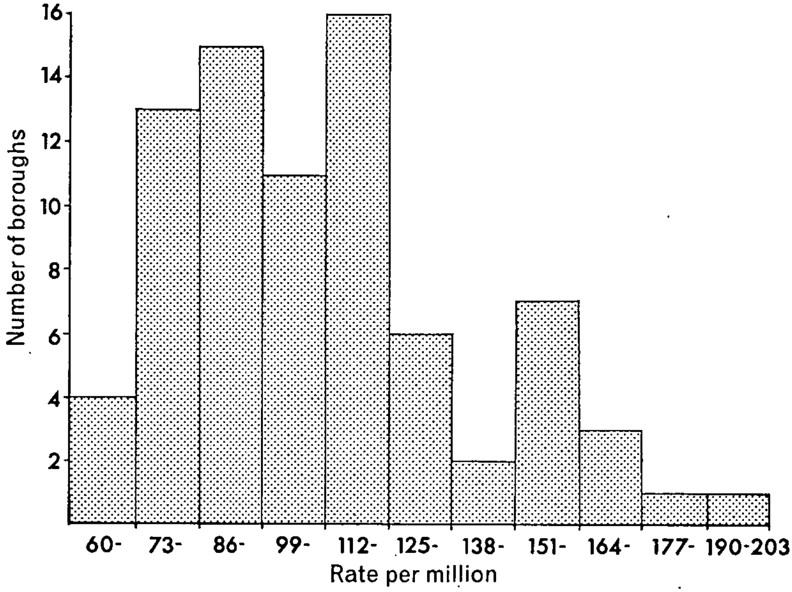


FIG. 9.7: The frequency distribution of the three-year average crude suicide rate per million, for seventy-nine county boroughs, 1950–2

The similarity between the coefficients could be taken as evidence for a prevailing standard of what a suicide is, if it were not for the absence of a common training for coroners. As they are frequently said to be succeeded by their sons or junior partners, there could be opportunity for indoctrination and this might account for the correspondence even if there were differences in standards between county boroughs.

The Medical Directory's lists of coroners show that 5 of the

TABLE 9.16. *Product-moment correlations between the three-year average suicide rates of county boroughs* 1950–2 and 1960–2*

	r	p
79 county boroughs	0·43	$< 0{\cdot}001$
39 county boroughs with the same coroner	0·45	$< 0{\cdot}01$
19 county boroughs with different coroners	0·49	$< 0{\cdot}05$

* Twenty-one boroughs were omitted because the coroner had changed during one of the six years from which the rates were obtained.

19 pairs of coroners shared the same address and were presumably partners and of these 5, 2 shared the same name and presumably were related. This proportion, even if they were thoroughly indoctrinated, is too small to account for the similarity of the correlations, and therefore the 19 must have similar standards to their predecessors which were not acquired by close contact.

Open verdicts

Deaths classified as open verdicts have not to our knowledge been systematically examined and reported in the literature, although their relationship to suicide has been the subject of comment and speculation. Of the deaths referred to the coroner, open verdicts are returned where there is uncertainty whether the correct judgement is death from natural causes, accident, murder, or suicide. A doubtful suicide may be categorized as an accidental death or an open verdict, depending on the standard of acceptable doubt held by each coroner. It would be convenient (although surprising) if there was a generally accepted standard as to what constituted this 'don't know' category.

We applied the same procedure to open verdict rates for coroners' districts. Table 9.17 shows that changing the coroner does alter the correlation coefficient. Although the difference between the two coefficients does not reach statistical significance at the 5 per cent level of probability, its size, particularly when compared with the finding for suicide rates, supports the view that coroners have differing standards for what constitutes an open verdict.

TABLE 9.17. *Product-moment correlations between the four-year average open verdict rates of coroners districts 1954–7 and 1964–7 (1)*

	r	p
55 coroners' districts	0·340	< 0·02
30 coroners' districts with same coroner	0·557	< 0·001
20 coroners' districts with different coroners	0·267	> 0·1

If the definition of suicide is standard but that for an open verdict is not, then the ratio between the two should vary con-

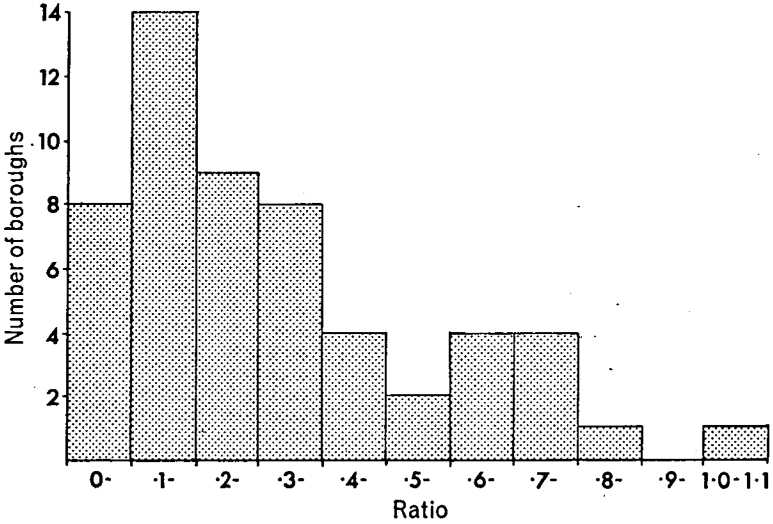


FIG. 9.8. The ratio of suicide verdicts to open verdicts for fifty-five county boroughs, 1955–7

siderably. Fig. 9·8 shows a range from 0 to 1·1. This means that in some areas there are no open verdicts and in others more open verdicts than suicide verdicts are recorded. Presumably, in areas with low ratios doubtful suicides are being categorized accidents, and in areas with high ratios doubtful accidents are being categorized as open verdicts; that is, the variation in open verdicts is not at the expense of categorizing suicides as open verdicts.

Demographic and geographic differences are also likely to contribute to this variation as they may influence the quality of the evidence available. For example, in the case of persons who were living alone and were found dead, it is likely that the absence of friends or relatives to supply evidence on state of mind contributes to uncertainty. Therefore areas with high proportions of one-person households are likely to have more open verdicts. Similarly, the well-known difficulty of deciding the correct verdict on a person found drowned means that county boroughs with rivers or those near the sea may have many open verdicts recorded.

We conclude from these findings that, in 19 county boroughs where the coroner changed, the incoming coroner used the same standard for a suicide verdict as his predecessor. As all 38 shared a similar training in the law and used the same set of instructions, namely, those set out in *Jervis on Coroners* (2), it seems a reasonable conclusion that they all used the same standard.

What is this standard? Presumably that stated in *Jervis on Coroners*—

> *Positive Evidence of Suicide*—to support a verdict of suicide there should be some actual evidence pointing to the event: the verdict should not rest upon surmise.

As nearly all coroners are lawyers and use *Jervis*, the likelihood is that this standard for suicide is in operation throughout England and Wales.

The same conclusion is not true for open verdicts. Nevertheless, open verdicts appear to be worth further inquiry, for the average ratio of about 2 to every 10 suicides gives some idea of the extent to which suicide may be under-reported through lack of evidence. The Registrar-General might help in this by reviewing the subject and account for them in his annual report in terms of age, sex, and causes of death.

REFERENCES

1. Unpublished Home Office data concerning numbers of open verdicts in Coroner Districts.

2. *Jervis on Coroners*, 9th edition (1957), p. 229 (London: Sweet and Maxwell).

SUBJECT INDEX

LIST OF AUTHORS CITED

Figures in italic refer to numbers in reference lists