

THE INTERNATIONAL PILOT STUDY
OF SCHIZOPHRENIA

REPORT OF
THE INTERNATIONAL
PILOT STUDY
OF SCHIZOPHRENIA

Volume 1

Results of the initial evaluation phase



WORLD HEALTH ORGANIZATION

GENEVA

1973

WHO Offset Publication No. 2

Volume 2 of the *Report of the International Pilot Study of Schizophrenia* will present the results of the two-year follow-up study.

A shortened version of Volume 1 of the report is in preparation for publication in the WHO *Public Health Papers* series in English, French, Russian and Spanish.

Reprinted 1974

WHO offset publications are intended to make generally available material that for economic, technical, or other reasons cannot be included in WHO's regular publications programme and would otherwise receive only limited distribution. They are usually reproduced by photo-offset from typescript, rather than by letterpress, and do not necessarily receive such detailed editorial revision as other WHO publications.

© World Health Organization 1973

Publications of the World Health Organization enjoy copyright protection in accordance with the provisions of Protocol 2 of the Universal Copyright Convention. For rights of reproduction or translation of WHO publications in part or *in toto*, application should be made to the Office of Publications and Translation, World Health Organization, Geneva, Switzerland. The World Health Organization welcomes such applications.

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Director-General of the World Health Organization concerning the legal status of any country or territory or of its authorities, or concerning the delimitation of its frontiers.

This report contains the collective views of the collaborating investigators and does not necessarily represent the decisions or the stated policy of the World Health Organization.

Cover design: Peter Davies

PRINTED IN SWITZERLAND

TABLE OF CONTENTS

Historical perspective	i
Authors' preface	iii
Chapter 1 Aims, scope and evolution of the International Pilot Study of Schizophrenia	1
Chapter 2 Diagnosis and distribution of schizophrenia	14
Chapter 3 Management and operation	39
Chapter 4 Description of the Field Research Centres	53
Chapter 5 Instruments	67
Chapter 6 Translation	86
Chapter 7 Units of Analysis	94
Chapter 8 Applicability and reliability of methods	112
Chapter 9 Characteristics of study population	147
Chapter 10 Psychopathological description of patients	158
Chapter 11 Clinical classification by computer	243
Chapter 12 Classification by cluster analysis	336
Chapter 13 A concordant group of schizophrenics	360
Chapter 14 Discussion	386
Chapter 15 Summary and conclusions	397
List of references	413

HISTORICAL PERSPECTIVE

In 1959 WHO convened an expert committee on the epidemiology of mental disorders (WHO, 1960). This committee reviewed the existing knowledge and stressed the need for reliable and valid data on the incidence and prevalence of mental disorders. The committee recommended that WHO should render assistance to activities concerned with psychiatric epidemiology in various countries of the world and coordinate and initiate research in this field. The committee felt that WHO should concentrate on problems which can be better solved through international coordination than by a single group and that it should explore the unique opportunities found in particular countries that require supplementation of the local effort. A series of studies were suggested which included studies aiming at a refinement of techniques of observation, classification, recording, and counting with regard to psychiatric disorder and the elucidation of problems of research design and studies of the influence of the sociocultural environment on the clinical condition and course of mental disorders. Other suggestions were made concerning studies on operational problems, such as the evaluation of psychiatric services and clinical research on problems of causation of psychiatric disorders.

The first steps to implement these recommendations were two important publications. One, by Dr D.D. Reid (1960) concentrated on epidemiological methods in the study of mental disorders. The second, by Dr T.Y. Lin and C.C. Standley (1962) focused on the scope of epidemiology in psychiatry.

Almost at the same time an informal meeting took place in Dr M. Kramer's office in NIMH and Drs S.W. Greenhouse, M. Katz, T.Y. Lin, B. Pasamanick and J. Zubin discussed the desirability and feasibility of studying the diagnostic process as a basis for developing effective methods for psychiatric epidemiology and cross cultural research.

A number of consultations and discussions followed that occasion until in 1964 WHO organized a Scientific Group meeting (WHO, 1964). This group, which was chaired by Dr R. Felix, recommended priorities for mental health research to WHO. The group put high priority on the development of methods necessary to carry out epidemiological research in a cross cultural setting. After the meeting of the Scientific Group in 1964, Dr Lin, in consultation with leading experts from several countries including Drs G.M. Carstairs, W. Caudill, E. Essen-Möller, R. Felix, M. Greenblatt, E. Gruenberg, M. Kramer, A. Lewis, E. Strömberg, J.K. Wing, L. Wynne and others, prepared the WHO meeting of investigators on comparative research on specific mental disorders in 1965. Discussion centred on WHO's research programmes in epidemiology of mental disorder and social psychiatry and an outline was produced for a long-term plan of studies in this area. Three basic papers were prepared for this meeting, one by Dr Lin, another by Dr Wing, and the third one by Dr Caudill. Consultations and work continued after this and several months later Drs Lin, Strömberg, Wing and Wynne worked out an initial plan of the IPSS which was presented to the Meeting of Investigators in the IPSS in 1966 (WHO, 1966). At the same time a grant was applied for and received from NIMH and

thus the funds necessary for the project were made available using three sources: WHO, NIMH and the collaborating centres. Soon after that the IPSS started.

The spirit of collaboration which was so very important in producing the initial proposals continued to be an essential factor in the further development of this study. Each important decision was reached after many consultations and many people made contributions at various stages of the project.

Some of the collaborating investigators and consultants are no longer connected with this project but their work and achievements were significant at the time when they were made and remain such today.

AUTHORS' PREFACE

The collaborating investigators have agreed that it would have been misleading to single out any one person as editor or principal author of this volume. Such a practice they felt would have been against the spirit of the project which from the earliest days has been a collaborative one. No effort has been made to distinguish between small and great contributions: rather they were all welcome. This spirit is well in line with the policy of WHO which has always been one of collaboration and coordination of efforts.

Each chapter of this book was drafted by someone delegated for the purpose. A list of chapters with names of the authors of the drafts are given below. In addition an editorial working group consisting of Drs T.Y. Lin, N. Sartorius, J. Strauss, E. Strömngren and J.K. Wing was established in 1969. This working group has made suggestions and comments about each of the drafts made by the various contributors who then redrafted their chapters and this was repeated several times until a pre-final draft was produced which was edited first by Dr Wing and then by the editorial working group. Dr R.W. Shapiro, Mr M. Kimura and Dr Sartorius gave the volume the final scrutiny and Mrs S. Shafner evened out the differences in style and improved the volume from the linguistic point of view.

Another important mechanism in the complicated process of producing this report was the circulation of its draft to all the collaborating investigators. For two drafts each of the collaborating investigators was requested to give detailed comments on a particular chapter. After that, on several occasions all the collaborating investigators commented on the complete drafts and again made many valuable suggestions.

On two occasions an entire draft was discussed by a full meeting of collaborating investigators: the third draft was discussed in February 1971 and the final draft in November of the same year.

The report was retyped several times and the secretaries of the Mental Health unit listed on page vi deserve our cordial thanks for their hard work, patience and endurance in this tedious and exacting task.

The chapters and the people who drafted them are listed below.

- Chapter 1: Aims, Scope, and Evolution of the International Pilot Study of Schizophrenia - Drs J.K. Wing and T.Y. Lin
- Chapter 2: Diagnosis and Distribution of Schizophrenia - Drs E. Strömngren and J.K. Wing
- Chapter 3: Management and Operation - Dr N. Sartorius and Miss E.M. Brooke
- Chapter 4: Description of the Field Research Centres. The original drafts produced by the Field Research Centres were condensed at Headquarters by Dr N. Sartorius, Miss E.M. Brooke with contributions by Dr R.W. Shapiro and Mr M. Kimura

- Chapter 5: Instruments - Dr N. Sartorius and Miss E.M. Brooke
- Chapter 6: Translation - Dr N. Sartorius and Miss E.M. Brooke
- Chapter 7: Units of Analysis - Dr N. Sartorius and Miss E.M. Brooke
- Chapter 8: Applicability and Reliability of Methods - Dr J. Strauss and Dr J. Bartko
- Chapter 9: Characteristics of Study Population - Dr N. Sartorius, Miss E.M. Brooke with substantial contributions by Dr R.W. Shapiro and Mr M. Kimura
- Chapter 10: Psychopathological Description of Patients - Dr N. Sartorius, Miss E.M. Brooke with substantial contributions by Dr R.W. Shapiro, Mr M. Kimura, Drs J. Bartko, M. Kramer and K. Williams
- Chapter 11: Clinical Classification by Computer - Dr J.K. Wing
- Chapter 12: Classification by Cluster Analysis - Dr J. Strauss with substantial contributions by Drs J. Bartko and W. Carpenter
- Chapter 13: A Concordant Group of Schizophrenics - Dr N. Sartorius, Dr R.W. Shapiro and Mr M. Kimura with contributions by Drs J. Bartko, M. Kramer and K. Williams
- Chapter 14: Discussion - Dr J.K. Wing
- Chapter 15: Summary and Conclusions - Dr N. Sartorius and Dr R. Shapiro with the contributions of Mr M. Kimura

The volume is thus in a real sense the product of many hands and the endeavour it records is the work of many more.

The Collaborating Investigators

COLLABORATING INVESTIGATORS

Principal Investigators

Dr T.Y. Lin (1965-1969)
Dr N. Sartorius (1969-present)

Other Headquarters' Investigators

Miss E.M. Brooke (1967-1970)
Dr F. Engelsmann (1967-1968)
Dr G. Ginsburg (1966-1967)
Mr M. Kimura (1971-present)
Dr A. Richman (1966-1967)
Dr N. Sartorius (1967-1968)
Dr R. Shapiro (1971-present)

Collaborating Investigators at the Field Research Centres

AARHUS: Dr E. Strömngren (Chief Collaborating Investigator) (1966-present)
 Dr A. Bertelsen (1970-present)
 Dr M. Fischer (1967-present)
 Dr C. Flach (1967-1969)
 Dr N. Juel-Nielsen (1967)

AGRA: Dr K.C. Dube (Chief Collaborating Investigator) (1966-present)
 Dr B.S. Yadav (1966-present)

CALI: Dr C. Leon (Chief Collaborating Investigator) (1966-present)
 Dr G. Calderon (1966-present)
 Dr E. Zambrano (1966-present)

IBADAN: Dr T.A. Lambo (Chief Collaborating Investigator) (1966-present)
 Dr T. Asuni (1966-present)

LONDON: Dr J.K. Wing (Chief Collaborating Investigator) (1966-present)
 Dr J. Birley (1966-1968)
 Dr J.P. Leff (1968-present)

MOSCOW: Dr R.A. Nadzharov (Chief Collaborating Investigator) (1966-present)
 Dr N.M. Zharikov (1966-1971)

PRAGUE: Dr L. Hanzlicek (Chief Collaborating Investigator) (1967-present)
 Dr C. Skoda (1968-present)

TAIPEI: Dr C.C. Chen (Chief Collaborating Investigator) (1966-1972)
Dr M.T. Tsuang (1966-1972)

WASHINGTON: Dr L. Wynne (Chief Collaborating Investigator, 1966-1967;
Collaborating Investigator, 1968-present)
Dr J. Strauss (Chief Collaborating Investigator, 1968-present;
Collaborating Investigator, 1966-1968)
Dr J. Bartko (1971-present)
Dr W. Carpenter (1968-present)

OTHER STAFF AND INDIVIDUALS WHO CONTRIBUTED TO THE IPSS

HEADQUARTERS (WHO, GENEVA):

Mental Health Unit

Various professional staff of the Unit since 1965, and the following:

Research Assistants: K. Barrett (1970-1972); A. Forster (1968-1970)

Secretaries: K.D. Ignoto (1969-present); S. Fischer (1972-present); V.M. Nicklin (1972-present); J. Mamboury (1967-present); S. Doyle (1972-present); M. Schwarz (1971-present); and many of the secretaries who previously worked in the Mental Health unit.

Health Statistical Methodology Unit: K. Uemura (1966-present); W. Gulbinat (1968-present)

Data Processing Unit: J.P. Bansal (1968-present)

Other WHO Units: Staff of the Budget, Documents, Finance, Personnel, Travel, and other units were very helpful at various times throughout the project.

NATIONAL INSTITUTE OF MENTAL HEALTH:

Project Officers:- Dr M. Kramer (1966-present); Dr H. Tuma (1967-1969)

Members of Advisory Group:- Dr M. Katz (1966-present); Dr L. Wienckowski (1968-present)

AARHUS:

Psychiatrist:- Dr N. Engkilde (1968-1969)

Psychiatric Social Worker:- L. Kann (1968-present)

Administrative, Secretarial and Other Staff:- J. Hildebrandt (1969-present); L. Aerø (1967-1969); G. Spencer (1969-present)

AGRA:

Psychiatrist:- Dr D. Parekh (1967-1969, 1970-1971)

Psychologist:- A. Kumar (1967-present)

Psychiatric Social Worker:- S.P. Gupta (1967-present)

Statisticians:- S.K. Handa (1967-1970); N. Kumar (1970-present)

Administrative, Secretarial and Other Staff:- O.P. Chauhan (1967-present)

CALI:

Psychiatrist:- Dr C. Córdoba (1968-1969)

Psychiatric Social Workers:- L. de Gambetta (1967-1969); R. de Guzman (1969-1970)

Translator:- D. de Zambrano (1967)

Administrative, Secretarial and Other Staff:- R. de Padilla (1968-1970); N. Ordoñez (1967-1970); Y. Otero (1967-1969)

IBADAN:

Psychiatrists:- Dr M.O. Akindede (1969-1971); Dr A.A. Marinho (1966-1967)
Dr M.O. Olatawura (1971-present); Dr C.O. Oshodi (1968-1969); Dr T. Otolorin (1968-1970)

Other Medical Officers:- Dr X. Fernandez (1969-1970); Dr V.O. Clairmonte (1969-1970)

Psychologist:- D.P. Dastoor (1966-1972)

Psychiatric Social Workers:- C.R.O. Barlow (1968-present); J. Ojesina (1966-1971)

Psychiatric Nursing Officers:- A.P. Coker (1966-present); A.O. Oyeneye (1968-present)

Translators:- J.T. Erinle (1966-1967); A. Johnson (1966-1967); J. Ojesina (1966-1967); E.O. Osanyin (1966-1967); A.O. Oyeneye (1966-1967)

Administrative, Secretarial and Other Staff:- J. Ogunniyi (1966-1967)

LONDON:

Psychiatrist: Dr R. Hirschfeld (1970-1971)

Statisticians:- P. Dugard (1971-1972); J. Nixon (1972-present)

Programmers:- C. Taylor (1966-1970); J. Gourlay (1966-present)

Administrative, Secretarial and Other Staff:- L. Astell (1969-1972)

MOSCOW:

Psychiatrists:- Dr V.G. Levit (1966-1972); Dr E.S. Petrova (1966-1972);
Dr A.N. Popova (1968); Dr M.S. Popova (1966-1968); Dr L.I. Teljuk
(1969)

Psychiatric Social Workers:- M.V. Ivanova (1966-1968); G.S. Kozlova (1969);
S.P. Matjushina (1968-1972)

Translators:- L.G. Krasjuk (1966-1972); V.N. Shalirina (1968-1972)

PRAGUE:

Psychiatrists:- Dr T. Dostal (1968-present); Dr M. Formánková (1968-present);
Dr E. Ledererová (1967-1968); Dr St. Růžicka (1968); Dr O. Vinar
(1967-1969)

Psychiatric Social Workers:- B. Hlouňová (1968-1969); M. Novotná (1969-
present); S. Pihrtová (1971); H. Svátaková (1969-1971)

Statisticians:- Dr T. Husák (1971); J. Janouch (1971)

Translators:- Dr T. Dostal (1968-present); Dr E. Ledererová (1967-1968);
M. Semotánová (1969); Dr C. Škoda (1968-present)

Administrative, Secretarial and Other Staff:- H. Jiroutová (1968-1971);
I. Kynclová (1969-1971); B. Prusíková (1968-1969); D. Vsetečková
(1971-present)

TAIPEI:

Psychiatrist:- Dr W.S. Tseng (1968-1972)

Psychologists:- M.Y. Wang (1968-1972); S.K. Yang (1967-1972)

Psychiatric Social Workers:- F.M. Chen (1967-1970); H.C. Chang (1970-1972);
M.L. Chiang (1966-1972); S.D. Yang (1969-1972); F.S. Yu (1967-1969)

Statistician:- Dr H.M. Chu (1967-1972)

WASHINGTON:

Psychologist:- J. Dent (1966-1969)

Psychiatric Social Workers:- R. Blatchley (1969); D. Davenport (1968);
S. Hirsch (1968-1969); Dr R. Savard (1968-1969)

Statistician:- Dr J. Bartko (1966-present)

Programmer:- K. Dorn (1969-1970)

Administrative, Secretarial and Other Staff:- R. Boesman (1969-present);
R. Silver (1967-1969)

CONSULTANTS

Drs G.M. Carstairs; W. Caudill; J.E. Cooper; J. Dent; E. Essen-Möller; R. Felix; J. Fleiss; M. Greenblatt; S. Greenhouse; E. Gruenberg; M. Hamilton; M. Katz; J. Klett; M. Kramer; A. Lewis; K. Rawnsley; A.V. Snezhnevskij; J. Zubin

PARTICIPANTS AT IPSS MEETINGS (UNLESS MENTIONED ABOVE)

Drs J. McFie; P. Schneider; E. Slater; J. Vaňá; J.M. Velasco-Alzaga and WHO staff members. In addition, exchanges of visits of collaborating investigators were held during the course of the study, and other staff of the involved field research centres and various experts from other institutions attended these meetings.

FUNDING

This project has been funded in roughly equal parts from three major sources - the Field Research Centres, the National Institute of Mental Health of the United States of America (Grant No. MH09239), and the World Health Organization.

The Field Research Centres provided professional, administrative and technical personnel, office space, supplies and in some cases facilities for data analysis. The centres often received valuable assistance from other institutions in their countries. The National Institute of Mental Health provided a grant for support of the study and the salaries of some staff members at headquarters and field research centres in developing countries were covered from this source. Funds were also utilized to cover the expenses of Exchanges of Visits of Collaborating Investigators, consultants and supplies. The World Health Organization contributed headquarters staff, part of the supplies and equipment, computer time and facilities for data processing and analysis and funds for the organization of meetings of investigators. The services of WHO units concerned with budget, finance, travel, reproduction of documents, and printing were another part of WHO's contribution.

AIMS, SCOPE AND EVOLUTION OF THE INTERNATIONAL PILOT STUDY OF SCHIZOPHRENIA

1.1 Epidemiological Programmes of WHO

The International Pilot Study of Schizophrenia (IPSS) is part of a long-term programme in epidemiological psychiatry developed by the Mental Health Unit of the World Health Organization. Plans for this programme stemmed from the recognition by WHO that epidemiological studies could play an important role in the establishment of programmes for the prevention and control of all diseases, a viewpoint shared by public health authorities in every country with a well-developed health programme. Adequate morbidity statistics were considered fundamental for the planning and evaluation of health services. These considerations were thought to apply as cogently to mental as to physical diseases and, in 1959, WHO set up an Expert Committee to consider the problems of psychiatric epidemiology. This Committee reported in 1960 (WHO, 1960) and, as a result, two monographs were published which provided an overview of the advantages and principal results up to that time of the application of the epidemiological method to psychiatric problems (Reid, 1960; Lin and Standley, 1962). A later outcome of the Committee's work was the monograph on mental health statistics (Kramer, 1969a). In 1964, the Scientific Group on Mental Health Research recommended that WHO give high priority to research activities in psychiatric epidemiology and social psychiatry (WHO, 1964).

The uses of epidemiology have been admirably stated by Morris (1964) as follows:

1. To assess changes over time in incidence, prevalence and mortality from diseases.
2. To carry out community diagnosis.
3. To assess the workings of the health services.
4. To estimate individual risks, on average, of acquiring various diseases and conditions.
5. To identify syndromes.
6. To complete the clinical picture and describe the natural history of chronic disease.
7. To provide clues to causes.

These headings are as relevant for psychiatry as they are for the rest of medicine. Even more relevant is Frost's statement in his classic paper on epidemiology (1927): "... since the description of the distribution of any disease in a population obviously requires that the disease must be recognized when it occurs, the development of epidemiology must follow and be limited by that of clinical diagnosis and of the rather complex machinery required for the systematic collection of morbidity and mortality statistics." In psychiatry, however, the problems of developing techniques that would enable clinicians to communicate with each other in a reliable way and to undertake meaningful comparative epidemiological studies appear to be immensely greater than in any other field of medicine.

In 1965, therefore, a long-term plan was drawn up in order to implement and give specific content to the general recommendations of the Scientific Group. This plan had four stages, of which the first two (Programmes A and B) were to run concurrently, while the third and fourth (Programmes C and D) were to depend to some extent upon the outcome of the earlier work (Lin, 1967).

Programme A was to be concerned with the standardization of psychiatric diagnosis, classification, and statistics, and was conceived as a long-term programme. Its aims were (a) to achieve a better understanding of the use of diagnostic terms by psychiatrists in different countries; (b) to facilitate the revision of Section V of the International Classification of Diseases scheduled for 1975; (c) to strengthen national programmes of mental health statistics; and (d) to foster the development of an international group of psychiatrists, biostatisticians, and epidemiologists, who could continue work in this field.

Twelve experts representing different schools of psychiatry and statistics were invited to participate during the first ten years of the project. This nuclear group was expanded at annual seminars by the addition of experts from the country hosting the seminar and neighbouring countries. At successive annual meetings the following topics were discussed by this group: functional psychoses, particularly schizophrenia (London, 1965); borderline psychoses, particularly reactive and psychogenic psychoses (Oslo, 1966); psychiatric disorders in childhood (Paris, 1967); psychiatric disorders in old age (Moscow, 1968); mental retardation (Washington, 1969); psychoneurosis and psychosomatic disorders (Basle, 1970); personality disorders and drug-dependence (Tokyo, 1971).

This programme has thus far been carried out as planned, using diagnostic exercises based on standard case histories and videotaped present state interviews with patients. The proposal for a new, triaxial classification of psychiatric disorders in children has recently been put to test use in several countries. A draft of an international glossary has been prepared and will be finalized in the near future. Some of the programme's results have already been published (Shepherd et al., 1968; Astrup and Odegaard, 1970; Rutter et al., 1969; Averbuch et al., 1970; Tarjan et al., 1972; see also reports of individual meetings, e.g. WHO 1970, 1971, 1973).

It was hoped that the deliberations of the nuclear group of 12 experts, together with the contributions of the other psychiatrists who participated

in the seminars, might be used as a basis for the next revision of Section V of the International Classification of Diseases, scheduled for 1975, and as a starting point for further practical work in this area. At a meeting held in 1972, the work completed up to that point was reviewed and plans for the future were adopted (WHO, 1973).

Programme B, entitled "Comparative Research on Specific Mental Disorders" was intended to determine whether comparable cases of mental disorder could be identified in various populations throughout the world (selected because they differed markedly in social and cultural characteristics). For such a programme, appropriate instruments would need to be developed for accurate and precise recording of the clinical and social information required. A team of research workers would also be needed in each of the areas under investigation and such teams would need to be trained. If this programme were successful, it would mean that the following objectives of the overall plan would have been attained: to establish whether certain specific mental illnesses were present in several culturally contrasting parts of the world; to develop systematic and reliable methods of recording symptomatic and socio-demographic data that could be used in comparable fashion by psychiatrists and social scientists from different schools of thought; to train research teams (particularly in the developing countries) in epidemiological methods; and to lay the foundation for true epidemiological studies.

Programme C was envisaged as developing out of Programmes A and B. It was thought that if the earlier programmes were successful, it would be possible to undertake proper epidemiological studies of specified mental illnesses in defined populations, using techniques and research teams developed during the earlier work.

Programme D was also planned to develop naturally from Programmes A and B. Its objective would be to devise and implement an international training programme in psychiatric epidemiology and social psychiatry.

These last two stages of the long-term plan are, of course, not yet operational, although preparations for initiating them are underway. Programme A, however, is well on the way to completion of its first phase, and in Programme B the IPSS eventually became the major vehicle of operation. The first results of this pilot study are published in the present volume.

1.2 Aims of the IPSS

In September 1965, a group of experts was convened in Geneva in order to consider how Programme B could best be implemented (WHO, 1965b). They had before them a memorandum setting out the long-term objectives of the WHO epidemiological programme and two working papers, one prepared by a psychiatrist and the other by a social scientist, suggesting that schizophrenia should be the main subject of the study and recommending the broad outlines of a study design.

It was thought that this design would be sufficient to answer many of the basic questions formulated in Programme B:

- a) In what sense can it be said that schizophrenic disorders exist in

different parts of the world? Do they differ in form or content? Does the clinical course differ?

b) Can other functional psychoses also be recognized and do they run a recognizably different course?

c) Can techniques be developed for recording and classifying symptomatology reliably?

d) Can teams of research workers be trained to use these techniques so that comparable observations can be made in both developed and developing countries?

These are not trivial questions. A consideration of the literature (see, for example, Chapter 2) illustrates the disadvantages of trying to solve problems concerning etiology and treatment without a prior demonstration that it is possible to agree on what condition is being investigated. The status of psychiatry within the expanding public health programmes being developed by WHO and national governments depends upon its having a solid clinical foundation on which planning and evaluation can be based. Moreover, it is only when these apparently simple questions have been answered that it will be possible to proceed to questions of etiology and therapy. It should be emphasized that this is true of all basic scientific work in psychiatry, not only of its social or cultural aspects, and certainly not only of its international aspects. Thus, the IPSS was addressing itself to problems which, if solved, could lead to the fruitful study of questions fundamental to the whole of psychiatry.

Moreover, it was felt that such a study would aid the development of a number of centres of psychiatric research, particularly in developing countries, which in time could come to serve as national and regional training centres and make their own epidemiological and cross-cultural contributions.

The factors favouring schizophrenia as the first subject of study were that there was a certain degree of agreement as to the chief features of at least a central group of disorders given this label; that numerous surveys had already been made and approximate incidence and prevalence rates established; that there was some evidence that the condition occurred at approximately the same rate in certain populations differing as widely as those of Bavaria, Bornholm, Baltimore, Taiwan, Japan, London, and Moscow; that almost the whole spectrum of psychopathology of the functional psychiatric disorders would be covered; and that the degree of severity and chronicity was such that in all societies schizophrenia was a personally crippling and socially damaging disease. In addition, there was an element of uncertainty to investigate, since certain studies had shown very high rates or very low rates of schizophrenia in isolated populations. Although the disease concept of schizophrenia had been challenged, no author had been able to produce as impressive an array of evidence in favour of any other approach or to show that an alternative concept would lead to a more useful way of studying what everyone agreed was a recognizable behavioural syndrome.

Because the case-finding procedures of the classical surveys had necessarily been imprecise, it was difficult to reach solid conclusions on the basis of comparisons between them. One of the major considerations, therefore, was that any study should adopt standard procedures both for collecting clinical information about each patient and for classifying the resulting data, so that a uniform diagnosis, comparable as between different areas, would be possible.

Since it did not appear feasible to examine sufficient samples of the general population to yield large enough groups of schizophrenic patients for comparative study (say, 100 from each area), the selection of a series had to be based on screening by psychiatric services. This selection method brought many difficulties in its train, not the least of which was that any idea of a proper epidemiological study (that is, a study based upon samples drawn in a specifiable way from the total population) had to be abandoned. This was not, however, an objective of the study in any case. The study was to be a pilot investigation with the main aim of discovering whether research teams could be trained to use specially developed techniques in a collaborative effort to find patients with various specified forms of mental illness, schizophrenia chief among them. The null hypothesis was adopted for test; that is, that examples of the main disorders under study would in fact be present in all areas and that no major differences would be found. Epidemiological work would come later (in Programme C), assuming that this first, limited, objective could be attained.

A corollary was that the research centres should be established in cities with reasonably well-developed psychiatric services and preferably with at least half a million inhabitants.

It was also decided that essential social and demographic data should be collected and a two-year followup be carried out in order to compare the course of the various psychiatric conditions in the different areas.

Following the decisions of this group of experts, further consultations and preparation took place, including the selection of centres, in order to work out the best means of implementation, and a draft design was considered by a meeting one year later at which most of the participating investigators were present. At this meeting plans were finalized and the IPSS came into being.

1.3 Design and Implementation

Basically, the design was that of a prospective followup study, with cases selected by a series of screening procedures and examined with standard instruments. The tasks were therefore to set up a Headquarters administration; choose the areas for study and establish working research teams in each one; devise the screening procedures appropriate to each area's services; select or create appropriate measuring instruments, translate them, and train the workers in their use; organize the regular despatch of data from the centres and its concurrent editing, coding, and analysis at Headquarters; and implement procedures for quality control, including meetings of all investigators, visits to each other's centres, training sessions,

feedback of results, and regular rounds by Headquarters staff.

The project was divided into three major phases:

- Phase 1: a preparatory phase, during which the technical and organizational groundwork would be undertaken both at Headquarters and in the centres.
- Phase 2: the main phase of the study, during which cases would be identified and data collected and sent to Headquarters.
- Phase 3: the followup phase, during which patients would be examined two years after their initial selection for inclusion in the series.

1.3.1 The first phase

Establishment of Headquarters. Headquarters was set up in the Mental Health Unit of WHO at Geneva, with a staff consisting of the principal investigator, a psychiatric epidemiologist, a social scientist, a statistician, a research assistant, and one secretary. The task of this team, together with their advisers, was to prepare the research instruments, draft the research procedures, select and train the collaborating investigators, and assist them in setting up their own organizations, keep control over the quality of data coming in, process and analyse the data, and convene meetings of investigators at which progress would be evaluated and further plans discussed.

Establishment of Field Research Centres. The nine centres were chosen according to the following criteria:

1. the existence of a network of services able to detect a substantial proportion of the likely cases of schizophrenia occurring in the population at risk (e.g., a first admission rate of 12 schizophrenic patients per 100,000 population per year);
2. the presence of several well-trained and motivated psychiatrists;
3. the possibility of setting up a simple reporting system so that potential cases would be known to the participating psychiatrists;
4. the recognition of a fairly distinct local culture or cultures;
5. the availability of census data covering the whole population;
6. the absence of very high death or emigration rates or a high prevalence of masking organic diseases that might make the diagnosis of schizophrenia difficult. (It was recognized that this criterion might be impossible to satisfy if the other conditions were to be met.)

The centres eventually chosen, after considerable travelling and much

debate, were situated in Aarhus (1); Agra (2); Cali (3); Ibadan (4); London (5); Moscow (6); Taipei (7); Washington (8); and Prague (9). These centres are referred to throughout the rest of this book as Field Research Centres (FRCs) and given the appropriate name or number. In fact, all the Centres were academic and/or research psychiatric centres and most of them served only an area in or near the city designated. Thus patients from the Washington Centre actually came from Prince Georges County, Md., an area adjacent to the Washington D.C. boundary. Details of catchment areas and other characteristics are given in Chapter 4.

In point of fact, the selection was based as much on the characteristics of the psychiatrists as on the characteristics of the Centres. It was necessary to find clinicians with training and experience who would appreciate the aims of the IPSS and the scientific values involved, and who had some personal experience with epidemiological work. All the collaborating investigators had at some time participated in scientific studies on schizophrenia. Their positions in their university hospital and research settings were such that they could make adequate resources available for this study. In addition, they represented some of the world's major contrasting cultures.

In each Centre two psychiatrists were designated as the collaborating investigators and, in some of the Centres in developing countries, funds were made available for a social scientist, research assistant, and clerical staff.

Admission criteria and screening. It was agreed that each series should include a sufficient number of young patients with functional psychoses of recent onset, covering the whole range of conditions including schizophrenia, mania, psychotic forms of depression and borderline psychoses. The lower age limit was set at 15 and the upper at 44, in both cases to avoid difficult nosological problems which, it was thought, should not be major concerns in a pilot project. Onset of the illness within 5 years of admission to the series was stipulated in order to exclude the difficult diagnostic problems associated with chronic forms of psychosis when acute symptomatology is not manifest. Certain behavioural and symptomatic criteria were also established as determining admission to the series.

All patients contacting the psychiatric services of each Centre would thus be screened to ensure that they satisfied the age and residence criteria, then to ascertain whether inclusion categories were present and exclusion categories absent (see below); in this case, they would be eligible for the series and detailed examination would commence.

Trial registration. Since several of the Centres were not familiar with procedures of the kind described above, it was decided that there should be a period of trial registration during which screening forms would be tested and the number of patients passing the various screens assessed. This would enable each participating centre to gain experience and to point out inadequacies in the procedure as practised in its own setting. On the basis of this trial, which proved very useful, a number of changes were introduced into the procedures eventually adopted for Phase 2. Some of the most relevant results of the trial registration period are given in Chapter 9.

Development of research instruments and training of collaborating investigators. The standard form of the Present State Examination (PSE) developed in the Medical Research Council Social Psychiatry Unit in London (Wing et al., 1967) was chosen as being best suited to the purposes of the IPSS. Certain modifications were introduced to adapt the PSE to the requirements of an international study, and the seventh edition of the PSE schedule was translated into the eight languages used by project investigators (see Chapter 6 for details). At the same time a start was made on assembling a glossary of definitions of the terms used. Preliminary versions of schedules for collecting demographic, social, and historical data were also prepared.

All the collaborating investigators were brought together for a week's training in the use of the PSE (see Chapter 5 for details) and for an initial assessment of inter-rater reliability.

As with registration, it was considered necessary for each Centre to undertake trial examinations in which all the schedules were to be tested on a group of cases. Each Centre was asked to select 26 patients - 12 with undoubted schizophrenia, 6 with doubtful schizophrenia, 6 with non-schizophrenic functional psychoses, and 2 with neurosis. In keeping with the criteria for admission to the main study, all trial patients were to be within the 15-44 age bracket and have no more than a five-year history of mental disorder. Many of the patients were interviewed twice by different examiners or rated by two clinicians at the same interview. The resulting schedules were sent to Headquarters for processing and analysis.

As a result of the trial registration and the trial examination of 26 patients, a substantial body of data was collected, together with a set of comments as to how the procedures and instruments might be made more useful. A full meeting of investigators was therefore convened in November of 1967 for a critical review of all the activities of Phase 1 and for agreement upon a definitive plan of operations for Phase 2. The changes made in the schedules are described in Chapter 5 and the results of the reliability study in Chapter 8.

1.3.2 The second phase

The plan of operation for Phase 2 was approved by the November 1967 meeting of investigators. All the instruments underwent revision and were reissued in their final form. The main aim of Phase 2 was for each Centre to collect at least 125 cases of functional psychosis, each documented in a standard manner. It was clear from the preliminary results that these cases would include a substantial number of patients with schizophrenia.

In order to identify patients to be included in the study, all patients contacting each of the FRCs were put through two screens, a demographic screen and a psychotic screen. The screens were designed to select patients with functional psychoses who would be likely to be available for follow-up for a period of two years from the time of their initial evaluation.

The Demographic Screen identified those patients having contacted each Centre during the course of the year from 1 April 1968 to 1 April 1969 who (a) had resided or slept regularly in the catchment area for the last six

months, and (b) were aged 15-44. This age range was chosen to exclude patients whose illness might be an early stage of presenile or senile psychosis at one end of the life-span, or childhood or juvenile schizophrenia at the other end. The residential requirement was designed to increase the likelihood of availability for followup.

The Psychotic Screen identified all of those patients who passed the Demographic Screen who did not fit any of the exclusion categories and who did fit at least one of the inclusion categories. Exclusion categories were chosen to screen out chronic patients and patients whose disorder may have been caused or significantly influenced by an organic condition. Since diagnostic practices vary, inclusion categories were symptoms rather than diagnostic labels.

The exclusion and inclusion categories appeared in their final form as follows:

Exclusion criteria:

(1) Severe psychotic symptoms in this episode probably present continuously for more than 3 years.

(2) Total hospitalization of 2 years or more in the last 5 years, including readmissions.

(3) Regular abuse of alcohol.

(4) Abuse of drugs acting on the C.N.S.

(5) Mental retardation with I.Q. estimated by psychiatrist to be less than about 70 before onset of present illness.

(6) Psychosis attributable to endocrine disorders (e.g., thyrotoxicosis, myxoedema, diabetes mellitus, or Cushing's syndrome).

(7) Psychosis attributable to metabolic or nutritional disorders, e.g., electrolyte disturbance, liver disease, vitamin deficiency.

(8) Evidence of acute or chronic brain syndrome, effects of brain surgery and other organic psychosis, not already specified in 6 or 7.

(9) Epilepsy.

(10) Severe hearing difficulties.) If serious enough to impede
administration of interview

(11) Severe difficulties in speech)
production or language (bad stammer,)
foreign dialect, etc.))

Inclusion criteria:

- (1) Delusions.
- (2) Definitely inappropriate and unusual behaviour.
- (3) Hallucinations.
- (4) Gross psychomotor disorder; over- or under-activity.
- (5) Social withdrawal.
- (6) Disorders of thinking, other than delusions.
- (7) Overwhelming fear.
- (8) Disorders of affect.
- (9) Depersonalization.
- (10) Self-neglect.

Inclusion criteria 1-4 automatically qualified the patient for inclusion, regardless of the severity of symptomatology. Categories 5-10 were considered as a basis for inclusion only if the symptomatology was present to a severe degree. In addition to these 10 criteria, provisions were made to allow the local psychiatrist to include a patient that he felt was definitely psychotic, even if he did not demonstrate any of the inclusion symptoms.

It was intended that Headquarters should monitor the age and sex distribution of cases in order to preserve a rough balance between Centres (see Chapter 9).

In addition to the 125 cases of functional psychosis, it was decided that 10 cases of neurotic depression should also be included in order to provide extra material for differential diagnosis.

Data collection. The collection of data would then proceed as follows:

(1) The mental status of the patient would be obtained during an interview conducted by a psychiatrist using the PSE. The present status would cover the patient's condition only at the time of interview and during the past month. The interview would take place within two weeks of the patient's contact with a psychiatric facility of an FRC.

(2) The past history of the patient and his illness would be obtained through interviewing of the patient or an informant by a psychiatrist, a psychologist, or a social worker using the Psychiatric History Form (PH).

(3) Social and demographic information on the patient and his family

would be obtained through interviewing of the patient or an informant by a social worker using the Social Description Form (SD).

(4) Physical and neurological examination performed by a physician would ideally be carried out on every patient, but in some centres where such an arrangement would be difficult examination could be limited to only those patients suspected of having central nervous system disorder of organic origin.

(5) After all the interviews, the psychiatrist would complete the Diagnostic Assessment Form (DA) to record the patient's diagnosis and prognosis, as well as the psychiatrist's reasons for his diagnostic judgement.

This material would then be sent at once to Headquarters for editing, processing and analysis.

Reliability. In order to provide a check on reliability, further examinations were to be conducted by two clinicians simultaneously or at a brief interval. Minimum requirements were specified as follows:

Five cases should be rated by two psychiatrists simultaneously, one acting as the interviewer and the other as observer-rater. An additional five patients should be interviewed twice by different psychiatrists with an interval of up to two weeks between the two interviews. If more than two psychiatrists are involved in any Centre, each pair should take part in ten such reliability exercises. The same principles should apply whenever a new psychiatrist joins the study.

At least one simultaneous interview should be conducted each month in order to discover whether any changes in the examination or rating characteristics occur in the course of the study. If more than two psychiatrists are involved, the number of reliability checks should be increased accordingly.

Videotapes and films should be made so that they can be rated by all the participating investigators during their annual meetings.

Simultaneous intra-centre ratings of past history and social description schedules should be carried out wherever possible, and a videotaped history interview should be made for subsequent rating by investigators from all Centres.

The results of these exercises are presented in Chapter 8.

Data processing and analysis. Headquarters accepted the responsibility for processing and analysing all the data collected from Centres. Data were checked for completeness and, in case of errors, corrected after consultation with the centre involved. The WHO Data Processing and Health Statistical Methodology Units provided invaluable assistance with the analysis. The collaborating investigators were solicited for suggestions concerning data analysis, and preliminary results were fed back to centres for discussion and reference. Centres with facilities and experts available for data analysis were consulted and considerable division of labour was effected in this way. Outside consultants were also used.

Meetings of investigators. Regular meetings of investigators were held in order to clarify and modify procedures, review results, discuss practical problems encountered in the field work, undertake reliability exercises, and generally assess progress and plan future activities. These meetings came to be regarded as an essential part of the study, invaluable for training, planning, and the maintenance of morale. International studies of this type cannot be conducted without them. Further details are given in Chapter 3.

One of the chief activities of later meetings was the planning of the followup phase of the study.

3.3.3 The third phase - followup

All patients in the study were intended to be followed up for a period of two years. In certain cases, a followup was also to be undertaken after one year in Centres where such exercises were not common practice in order to test out necessary procedures. The existing schedules needed adaptation in many cases to fit the circumstances of the followup (particularly the Followup Psychiatric History Form and the Followup Social Description Form) and these revised schedules needed to be tested. Monthly reliability testing was planned as before.

3.4 Chronology of the IPSS

A brief chronological account of the major events taking place during the three phases is given below in order to summarize the activities and provide an overview of the whole operation.

It will be remembered that Programme B was planned during 1965 as part of the overall epidemiological programme of WHO. The IPSS was adopted as the major vehicle of this programme in September 1965 (WHO, 1965b). Preliminary discussions concerning the selection of Centres and collaborating investigators took place between October 1965 and February 1966, and the preparation of research instruments continued during this time up to June 1966. The first meeting of collaborating investigators, which occurred in July 1966, approved in principle the procedures to be adopted during Phase 1.

Phase 1 (August 1966 - November 1967)

- (1) Improvement of instruments for use in Phase 1 (August 1966 - April 1967).
- (2) Selection and establishment of Centres (January - November 1967).
- (3) Translation of research instruments into local languages of Centres (January - April 1967).
- (4) Trial registration (March - April 1967).
- (5) Training of collaborating investigators in the uniform application

of research instruments and procedures (May 1967).

(6) Trial examination and rating of 26 patients (July - August 1967).

(7) Assessment of Phase 1 and agreement on procedures of Phase 2 (November 1967).

Phase 2 (November 1967 - June 1969)

(1) Revision and finalization of research instruments and procedures for use in Phase 2 (November 1967 - March 1968).

(2) Identification and collection of data on at least 125 patients from each FRC (April 1968 - September 1969).

(3) Preliminary assessment of the results of the initial stage of Phase 2, and pre-final draft of Followup PH and SD forms (July 1968).

(4) Assessment of results of Phase 2 and finalization of plan of operation and instruments for followup study (May - June 1969).

(5) Data processing and analysis (January 1969 - October 1971).

Followup phase (June 1969 - October 1971)

(1) First year followup of patients (September 1969 - March 1970).

(2) Continuation of data processing and analysis.

(3) Further assessment of the results of Phase 2 (December 1969).

(4) Second year followup study (March 1970 - October 1971).

(5) Finalization of draft report, Volume I (January - December 1971).

Data analysis for the whole project will be completed in 1974, and Volume II of the report will contain further results from Phase 2 as well as the results of the followup study.

DIAGNOSIS AND DISTRIBUTION OF SCHIZOPHRENIA

2.1 Development of the Concept of Schizophrenia

In any discussion on the concept of schizophrenia the names of Emil Kraepelin and Eugen Bleuler are invariably mentioned. There are very good reasons for this. First, any person seriously engaged in studying the problem of schizophrenia has some knowledge of the main ideas of Kraepelin and Bleuler. Second, it is difficult to see that subsequent workers have added anything of comparable acceptability to the understanding of schizophrenia. Thus, it is natural to begin a discussion of the concept of schizophrenia with a recapitulation of Kraepelin's and Bleuler's teachings. Certain misunderstandings concerning the viewpoints of these two classical conceptualists are very widespread and some of them have given rise to highly emotional attitudes, especially antagonism to Kraepelin.

It is generally assumed that Kraepelin's concept of dementia praecox was first presented on the occasion of a meeting that took place on the 27th of November 1898 at the University Psychiatric Clinic in Heidelberg, which he headed at that time. He called his historic lecture, "Zur Diagnose und Prognose der Dementia praecox". The 1893 edition of his textbook had already used the term dementia praecox, but at that time it covered a more narrow concept, excluding "Katatonie" and "Dementia paranoides". In the 1896 edition Kraepelin grouped these three forms together under the sub-heading "Verblödungsprozesse" (deterioration processes) within the main group "Stoffwechselkrankheiten" (metabolic disorders), along with myxoedema and cretinism. What happened in the 1898 lecture was essentially that Kraepelin stressed the importance of distinguishing dementia praecox from manic-depressive disorder, because the prognosis was so different in these two kinds of psychosis. His lecture was, however, partly misunderstood, probably because he paid too much attention to the difficulties of distinguishing the two diseases. The majority of those who took part in the discussion could not see that he had actually added greatly to the clarification of the concept of dementia praecox. This unpleasant discussion probably taught Kraepelin a lesson. In the edition of his textbook appearing in 1899, Kraepelin presented his views on dementia praecox clearly in the form in which we know them today, with the main point being the correlation between symptomatology and prognosis. In this connection, however, a common misconception should be corrected. In contrast to what is nearly always assumed, Kraepelin did not regard dementia praecox as having an invariably poor prognosis. His statistics had shown him that no fewer than 13 per cent of his cases seemed to recover completely. Conscientiously, he admitted that there might nevertheless be some minor traces of the disease in these cases, but it is obvious that he did not regard recovery as incompatible with the diagnosis of dementia praecox.

Kraepelin's dementia praecox, which was now a separate main group, in-

cluded three types: hebephrenia, catatonia, and paranoid deterioration. Regardless of the type, the following symptoms were regarded by Kraepelin as characteristic of dementia praecox: hallucinations, usually of an auditory or tactile nature; decrease of attention towards the outer world; lack of curiosity; disorders of thought, especially of the "Zerfahrenheit"-type with unusual and partly incomprehensible associations; changes of speech consequent upon the thought disorder, such as incoherence; lack of insight and judgement; delusions; emotional blunting; negativism; stereotypes. Kraepelin stressed the important point that these symptoms were present in spite of clear consciousness and unimpaired perception and memory, disorders of consciousness occurring only in certain rare cases of acute excitement.

Kraepelin was an empirically-minded clinician who felt it his first duty simply to describe what he had observed. His great contribution lay in the description of the symptomatology and the course of a great number of psychiatric cases. He was very reluctant to conceptualize his observations. He felt that nothing solid could be said about the etiology of dementia praecox and therefore did not spend much time on speculations of this sort. Occasionally he mentioned that he believed the etiology to be of a somatic, probably metabolic, nature, but he stressed at the same time that there was no evidence to support this viewpoint.

The contribution of Eugen Bleuler to the knowledge of schizophrenia did not in any way conflict with that of Kraepelin but was a most important addition to it. In the preface to his classic work, "Dementia praecox oder die Gruppe der Schizophrenien", which was completed in 1908 but not published until 1911, he stated that "the whole idea of dementia praecox comes from Kraepelin; also the grouping and the description of the symptoms is practically due only to him", and he added concerning his own contribution, "an important part of the attempt to expand the pathology simply consists in the application of Freud's ideas to dementia praecox".

In addition, Bleuler is generally regarded as having initiated an extension of the concept of dementia praecox, including — in supposed contrast to Kraepelin — cases with a favourable outcome. As far as his classic monograph is concerned, the opposite is actually true. Whereas Kraepelin, as mentioned previously, had found that a considerable number of his cases recovered, Bleuler in his introductory definition of schizophrenia says that this term is used to label a group of psychoses, the course of which may vary and may, at any stage, be arrested or even reversed, but which probably never reaches *restitutio ad integrum*. He adds, "they are characterized by a specific alteration of thinking and emotionality and of the relations to the surrounding world, an alteration that does not appear anywhere else".

Bleuler's well known innovation consisted mainly in his attempt to separate primary symptoms ("Grundsymptome"), caused directly by the etiological agent, from secondary symptoms ("akzessorische Symptome"), which represented the psychological reactions of the personality to the changes of a primary nature. He stressed that in many cases the secondary symptoms might be predominant and that most of the psychological content of the psychosis might be interpreted as psychologically determined and very often be understandable only along psychoanalytic conceptual lines. The primary

symptoms were regarded as characteristic of schizophrenia and supposed to be present to some degree during all stages of the course of the illness (Dauersymptome), whereas the secondary symptoms were nonspecific, occurring frequently also in non-schizophrenic disorders. Primary symptoms were the following: disturbance of associations, thought disorder, changes in emotional reactions, tendency to prefer fantasy to reality, tendency to seclude oneself from reality, being autistic. The secondary symptoms were hallucinations, delusions, catatonic symptoms, and all kinds of behavioural anomalies.

Thus, the foundations of the concept of schizophrenia were laid by Kraepelin and Bleuler. Much detailed knowledge concerning schizophrenia has been added since then but no major changes have been made in the concept, although there have been some modifications of interest. One of these is based on the work of Meyer and Sullivan, both psychiatrists who contributed to the broader conceptualization of schizophrenia that is often used in the United States. Meyer emphasized the view that "dementia praecox" was a collection of habit patterns, exaggerations of normal behaviour, often developed in response to continuing environmental stress. He disagreed with the concept that "dementia praecox" was a discrete disease entity and placed less emphasis on the importance of prognosis. Sullivan extended this view by stressing the importance of disturbed interpersonal relations for the etiology, diagnosis, and treatment of schizophrenia. Although Sullivan stated that there was also a disorder, "dementia praecox", that probably was organically determined, this part of his teachings has received less emphasis.

To summarize present knowledge and thinking concerning schizophrenia, the formulation made in recent years by Manfred Bleuler may be of value. On the basis of his own clinical experience, his research and the studies of his many collaborators, and his profound knowledge of the world literature of schizophrenia, Bleuler has attempted to point out what he regards as the essence of this disease (1965, 1972).

First of all, Bleuler stresses that schizophrenia is a psychosis. Without psychosis, there is no schizophrenia. This does not mean, however, that every schizophrenic should be overtly psychotic at every stage of the disease.

Next, Bleuler points out which kinds of psychosis are definitely not schizophrenic. Psychoses characterized by impaired memory and deterioration of the intellect are not schizophrenic. Neither are psychoses in which consciousness is disturbed, psychoses characterized by quantitative changes in the emotions, nor psychoses that arise in direct connection with physical illness. The same applies to psychoses that develop immediately following mental traumata, with rapid recovery when the traumata are removed.

Bleuler feels that nowadays it must be recognized that schizophrenia can end in complete recovery. Although in many cases — as stressed by Langfeldt — schizophreniform psychoses with favourable outcome can be distinguished at an early stage from true schizophrenias, this distinction cannot always be made on a symptomatological basis.

As positive psychopathological characteristics of schizophrenias, Manfred Bleuler has named the following: the peculiar combination of deeply psychotic

and quite normal mental processes within the same personality; splitting of associations; changes in emotionality; and experiences that the patient's mind or personality are being influenced or controlled by other agencies.

In addition to these specific traits there are a number of criteria that have a statistical association with schizophrenia but are in no way necessary for the diagnosis of schizophrenia: schizoid premorbid personality; rarity of pyknic body type; and overrepresentation of schizophrenics and schizoids among the patient's relatives.

With regard to etiology and pathogenesis, Bleuler believes that certain kinds of conflict, primarily between the patient and his close relatives, are of great importance for the development of schizophrenia. Such conflicts do not, however, lead to schizophrenia in all those who are exposed to them; a special psychological predisposition is necessary. Genetic factors may in part be responsible for this predisposition.

It is obvious from this formulation that Manfred Bleuler's concept of schizophrenia is, in the main, a symptomatological one.

Why is a concept of schizophrenia necessary at all? Firstly, because we have the term. The word schizophrenia has come into such widespread use that it is necessary to have a practical definition of it in order to keep public discussion of schizophrenia within reasonable limits. Nowadays it is not enough that psychiatrists, through their acquaintance with hundreds or perhaps thousands of schizophrenic patients, have acquired a strong feeling of what schizophrenia is, or that this feeling makes it possible for them to exchange ideas with colleagues without paying too much attention to sharp definitions (although this would probably be the most scientific, unbiased attitude at the present stage of research when nothing definite is yet known about what lies behind schizophrenia). For the benefit of non-professional contemporaries who enjoy talking about schizophrenia without knowing what it is, we must introduce some simple definitions of the concept.

We can readily agree that an etiologically based concept cannot be established today. Something is known about etiology but certainly not enough. We know that genetic factors play some role but also that other etiological factors must be involved. We have no idea whether the genetic factors are the same in all cases, or whether from a genetic viewpoint there are many schizophrenias or perhaps even polygenic determination of schizophrenia. With regard to exogenous factors, some are known to play a certain role, but it is very difficult to find anything that these factors have in common, since they are of very different types. Kraepelin's attempt to create a nosological concept of dementia praecox was unsuccessful, as all later attempts in this direction have been. What gives so many experienced psychiatrists the feeling that the cases they diagnose as schizophrenia have something essential in common has nothing to do with etiology or outcome but only with symptomatology. It may well be that there are many kinds of etiology. There are certainly great variations in outcome, even though some psychiatrists still maintain that complete recovery is impossible. What is certain, however, is that in the great majority of cases the course and outcome depend, to a very high degree, on the kind of therapy and care given to the patient.

To this it may be added that when a schizophrenia-like picture develops

in direct connection with a physical disorder or acute mental stress, there are in most cases some symptoms that are not characteristic of schizophrenia: in the physiogenic disorders, disturbances of consciousness and some degree of amnestic syndrome; in the situationally determined cases, strong emotional reactions, concentration of thoughts on the acute conflict, and very often also disturbances of consciousness. If such cases are excluded as not being characteristically schizophrenic, we are left with a large group of schizophrenias in which we cannot find any trace of physical disorder or any acute psychological stress that could explain why the schizophrenia developed at this particular point in time. These facts are, to a certain degree, responsible for the feeling of so many psychiatrists that genetic factors must be mainly responsible for the development of schizophrenia. On the other hand, twin studies show clearly that exogenous factors are of great importance in the etiology of schizophrenia.

In his recent papers, Manfred Bleuler has formulated an etiological theory of these cryptogenetic schizophrenias. He is certain that genetic factors are at work but feels that they may not be pathogenic per se. What is pathogenic is a disharmonious combination of genes. These combinations cause disharmonies of character in the sense that the personality comprises tendencies that are incompatible with one another. Under psychological stress, especially of the type that occurs within the close group of relatives, this incompatibility may lead to schizophrenic breakdown. There are, however, no specific stresses at any specific period of development that cause these maladaptations. The etiology is heterogeneous with regard to the exogenous factors as well as with regard to the genetic factors.

In his concept of schizophrenia M. Bleuler stresses the dissociation of personality. What is characteristic of the schizophrenic patient is the coexistence of incompatible components. These components may not be of a definitely pathological nature. Most "schizophrenic" mechanisms occur in some form also in normal persons; in day dreams as well as in dreams that occur during sleep, the changes in personality, the unrealistic ideas, and the hallucinations are identical in nature with much of schizophrenic experience. What makes a person schizophrenic is that this content is also present when he is awake and should be able to repress the unrealistic content as he compares it against reality. His disease thus consists in his inability to correct his fantasies by contact with reality. Subjectively, he therefore has two realities that he does not allow to come into conflict with each other. Such dissociation processes are essential to the concept of schizophrenia.

With respect to the problem of the "primary" or pathognomonic symptoms of schizophrenia, one focus of interest in recent years has been the symptom of autism. Within many psychiatric schools there is an inclination to say that, if any symptoms are pathognomonic of schizophrenia, autism must be among them. This has given rise to terminological confusion because Eugen Bleuler in some of his works tended to use the term "autism" not only for certain mechanisms seen in schizophrenic patients but also for a more general attitude that may be present even in normals. His monograph on "Das autistisch-undisziplinierte Denken in der Medizin und seine Überwindung" is a famous expression of this viewpoint. The difference between this wider

concept of autism and the autism characteristic of schizophrenia is obviously that while the normal person may experience all kinds of wishful "autistic" thinking, he will also be able to correct the wishful aberrations, at least to himself. The schizophrenic patient, however, is unable to see the conflict between his two realities, of which one is autistic.

The term "autism" has also been applied to certain kinds of abnormality in early childhood but, as Bosch (1962, 1970) has pointed out, Bleuler's concept cannot properly be applied to these children; and many workers consider "early childhood autism" to have no connection with schizophrenia (Rutter, 1972).

"Autism" and "dissociation of personality" are different expressions of the same mechanism. However, before this can be accepted as the core of the classical concept of schizophrenia, it is necessary to discuss hysteria, in which dissociative mechanisms are common, and certain personality disorders in which the affected individual may be so unable to differentiate between fantasy and reality that he actually acts out his fantasies in real life, sometimes with appalling consequences. The differences between such individuals and schizophrenics are fundamental. In true schizophrenic autism the inner life acquires definite autonomy, the emotions being detached from the real surroundings. In the other group of individuals who present dissociative phenomena, it is always evident and even conspicuous that, however self-centred they are, their emotions depend on the reactions of the environment; they are dependent to an extreme degree on the human beings surrounding them and in this sense are even less "autistic" than normal individuals.

It is perhaps in an effort to avoid such intricate problems that some psychiatric schools have attempted to define schizophrenia in purely symptomatological terms. Kurt Schneider's distinction between "first rank" symptoms and other symptoms may be taken as a fruitful example of this approach. Schneider points out that this distinction does not aim at singling out symptoms that are supposedly direct expressions of some basic schizophrenic "disorder". The first rank symptoms are of no theoretical value, they are not regarded as "primary" symptoms, "basic symptoms", etc.; they are simply symptoms that empirically have turned out to be of the greatest diagnostic importance. First rank symptoms are: "Gedankenlautwerden" (patient hearing his own thoughts spoken aloud); hearing voices talking to each other, voices that comment on the behaviour of the patient; feelings of influences on bodily functions; interference with thoughts; thought stealing; communication of own thoughts to others; and feelings of being influenced from the outside with regard to emotions, drives, and volition. When these symptoms are present and no physical disorder is to be found, Schneider regards it justified to make the diagnosis of schizophrenia. He points out, however, that the same symptoms may occur in diseases with a verified somatic etiology, such as alcoholic or epileptic psychoses.

Another important consideration in the selection of these first rank symptoms, according to Schneider, is that they should be reasonably easy to ascertain. Second rank symptoms are other hallucinations than those mentioned; paranoid ideas; perplexity; depression or elation; feeling of loss of emotions. First rank symptoms may not always be present and are not absolutely necessary for the diagnosis of schizophrenia, but for the differential

diagnosis between schizophrenia and manic-depressive disorder they are usually of decisive importance.

If successful, Schneider's approach would make it possible to define a central group of patients who present with such symptoms, in the absence of dysmnesic features or disturbance of consciousness. It has been suggested by many that such a group would tend to have a poorer prognosis than any group of patients without the "nuclear" features but defined in other ways.

Such attempts at contrasting nuclear and peripheral schizophrenias, or process and reactive schizophrenias, have taken many different terminological forms. Langfeldt, for instance, distinguishes between schizophrenia and "schizophreniform states". He feels that the introduction of the concept of schizophrenia has led to an unfortunate grouping of a number of very different disorders into a heterogeneous category. He would prefer to adhere to the concept of dementia praecox as defined by Kraepelin and to apply the term schizophreniform states to the extra kinds of disorders that have gradually been included under the label of schizophrenia. He adds that, on the whole, the schizophreniform psychoses have a good prognosis, that they arise frequently in immediate connection with mental stress, and that, in the acute states, they are often accompanied by disturbance of consciousness. Genetically, he considers them unrelated to the nuclear schizophrenias.

In this context it may be apposite to mention the schizo-affective psychoses. When Kasanin (1933) coined this term he applied it to cases "characterized by a very sudden onset in a setting of marked emotional turmoil with a distortion of the outside world.... The psychosis lasts a few weeks to a few months and is followed by a recovery". Vaillant (1962, 1964) is perfectly right in pointing out that Kasanin's cases appear indistinguishable from many of the schizophreniform psychoses of Langfeldt. Vaillant also draws attention to the terminological confusion that has arisen concerning the schizo-affective psychoses; in the official American Psychiatric Association nomenclature they have been described as a subgroup of schizophrenia showing significant admixtures of schizophrenic and affective reactions; "on prolonged observation, such cases usually prove to be basically schizophrenic in nature". This obviously is a fundamental departure from Kasanin's definition.

The attempts to describe a nuclear group of schizophrenia have their complement in attempts to identify groups of patients who suffer from schizophrenia but who do not show recognizable nuclear symptoms. In many cases of "simple" schizophrenia there may for long periods be a complete absence of first rank symptoms, the diagnosis being mainly based on the gradual deterioration of personality in the direction of autism and emotional blunting. Occasional short-lasting, mild, but clear-cut psychotic episodes help to confirm the diagnosis of schizophrenia. A similar course may characterize Hoch and Polatin's "pseudo-neurotic schizophrenia", in which definite psychotic symptoms of a schizophrenic nature occur for brief periods, often at long intervals, while the rest of the picture is characterized by symptoms that at first sight seem neurotic, such as anxiety, hypochondria, and anancastic mechanisms; on closer scrutiny they are usually distinguishable from neurotic phenomena because of their lack of comprehensibility or of secondary gain. Only a minority of these cases proceed in the direction of more classi-

cal schizophrenia. Similarly, it may be difficult to ascertain the schizophrenia behind the picture of "pseudo-psychopathic schizophrenia" in the sense of Dunaif & Hoch (1955). On the surface these patients seem to be characterized by antisocial acting-out behaviour, often with sexual deviation and offences, stemming from a disorganized sexuality.

Whenever a wide concept of schizophrenia is applied, one can easily single out special types that seem to differ essentially from nuclear schizophrenia with regard to symptomatology, etiology, and course. In contrast, attempts at subdividing the nuclear group have not been very successful. It appears as though the differences between the classical subgroups of simple, hebephrenic, catatonic, and paranoid schizophrenia are purely symptomatological, the differences being partly of a pathoplastic nature, determined by age at onset of the disease, and by cultural and other environmental factors, not the least of which is therapy. Similarly, for most psychiatrists it has been difficult to find empirical support for the elaborate subdivisions of schizophrenia created by Kleist and his followers, primarily Leonhard.

In the systematization of schizophrenia developed at the Institute of Psychiatry of the Academy of Medical Sciences of the USSR, the basic criterion for differentiation within the group of schizophrenias is the character of the course. Three main forms are distinguished, depending on whether the course is continuous, recurrent, or "mixed" ("attack-like" and progressive): 1) continuous schizophrenia, with subtypes (a) sluggish schizophrenia, (b) progressive (paranoid) schizophrenia, (c) malignantly developing schizophrenia; 2) periodic schizophrenia; 3) shift-like progressive schizophrenia, with subtypes (a) mild progressive type, (b) progressive type, (c) malignant type. It is felt that these forms differ from each other in terms of symptomatology, development, and response to treatment, and that there are significant differences in their pathogenesis.

In conclusion, it seems natural to attempt to analyse the reasons for these wide discrepancies in the use of the term schizophrenia. It is reasonable to assume that the majority of those applying the term would prefer to use it for a group of conditions with similar etiology, symptomatology, course, and response to therapy. The problems arise mainly because some of these ingredients are unknown. At the time when treatment for schizophrenic conditions was practically nonexistent and knowledge of etiological factors very scarce, the most important task of the psychiatrist was to attempt to indicate the prognosis. This he could only do on the basis of knowledge of the previous course and symptomatology. Kraepelin was the founder of this line of empirical research. He and his followers also hoped, of course, that this technique would lead to nosological differentiation. Later, however, when psychiatrists came to adopt a particular theory of etiology they tended to allow this to influence their diagnostic principles: whenever an etiologic factor was "found", a diagnosis would be made, regardless of symptomatology and course. This applied to any kind of etiological theory, be it psychodynamic, genetic, or other.

Progress in therapy has also seemed to give rise to new distinctions. For example, the application of convulsive therapies during the 1930s yielded one important result: the distinction between those "schizophrenic"

states that responded to therapy and those that did not. The separate classification of "schizophreniform states" that were supposed to be etiologically different from schizophrenia was accepted by many. Later, during the first phase of pharmacological treatment it turned out that many drugs had some therapeutic influence on nearly all kinds of psychiatric disorder. This led to less attention being paid to diagnosis; it became fashionable to speak of "anti-psychotic" drugs and, in a way, to return to the "Einheitspsychose" of more than a hundred years ago, with the modification that the term schizophrenia was now used for all these psychoses. New developments in drug therapy have again led to differentiation: the specific action of anti-depressives and of lithium on affective disorders have made it natural to regard these as something fundamentally different from schizophrenia.

Since a sufficient understanding of etiology is still lacking, no nosological concept can at the present time be regarded as a solid basis for international agreement on the delimitation of schizophrenia. Likewise, knowledge about premorbid personality and previous history is often difficult to ascertain in many cultural settings. Reaction to therapy is unknown in the first stage of the disease. It would seem, therefore, that the only type of concept of schizophrenia that can be applied by psychiatrists in all cultural settings is a mainly symptomatological concept. A standardized technique for examining the present state seems to be the most reasonable tool for this purpose. The main merit of this instrument is that results obtained with it would be reproducible. It would then be possible to discover whether schizophrenia defined in a standard way, in terms of the data collected using the instrument, has any relationship to etiology, course, or response to treatment. That would be a matter for empirical investigation.

2.2 Recent Discussions concerning the Boundaries of Schizophrenia

It is clear from what has been said in the previous section that there have long been opposing trends in views concerning the nature of schizophrenia. Kraepelin's great simplification brought unity out of diversity by labelling several overlapping symptom-clusters with the same generic name, justifying this procedure on the basis of a common course and outcome (and, as Kraepelin thought, a common cause).

Eugen Bleuler's work appeared to strengthen the foundations of this unity by specifying the fundamental symptoms that had to be present, whatever the form or content of the rest of the clinical picture, if the disease "schizophrenia" were to be diagnosed. However, these fundamental symptoms were themselves so insusceptible of definition that, although a disease entity had been created, its boundaries could be made to vary within wide limits by those making the diagnosis. Thus diversity re-entered through the door that should have shut it out. The majority of contemporary discussions focus on the central question of how much can be considered to lie within the limits of the disease; and most controversy turns on the place of latent, borderline, pseudoneurotic, pseudopsychopathic, or other non-Kraepelinian forms, which can only be regarded as schizophrenic in symptomatic

terms by appeal to one or other of Bleuler's fundamental symptoms — "disturbances of association and affectivity, the predilection for fantasy as against reality, and the inclination to divorce oneself from reality." There is a world of difference between such "fundamental" symptoms, which shade insensibly into normality and can be present when none of the Kraepelinian syndromes can be detected, and the "first-rank" symptoms of Schneider — voices discussing the patient in the third person, thought broadcast and so on — that Bleuler would have regarded as accessory or secondary.

The only way out of such dilemmas is empirical — to remember that disease concepts are created to be useful. If they are not useful they should be discarded. The important question then becomes whether a given delimitation of the concept of "schizophrenia" can be understood and uniformly applied by clinicians and scientists and, if so, whether it gives them helpful clues concerning causes, treatments, course, or eventual outcome. Such constraints on concept-building are now more widely appreciated and this volume is itself evidence that psychiatrists recognize the value of critically reexamining the current diagnostic systems.

Most of the recent work can be seen as an attempt, not always deliberate or consistent, to deal with the problems imposed by this polarization in the various concepts of schizophrenia. An excellent summary of the current position may be found in a book edited by Rosenthal & Kety (1968), which contains statements by protagonists of several of the principal schools of thought. Empirical support for retaining a broad spectrum of disorders under the general rubric of schizophrenia comes from the finding, in recent genetic studies (Kringlen, 1967; Tienari, 1963; Fischer, et al., 1969; Gottesman and Shields, 1966), that the monozygous co-twin of a "schizophrenic" patient is often suffering not from a readily recognized form of the disease but from some other disorder or disorders, variously labelled "borderline", "schizoid", or "psychotic". Shields refers to these conditions as "schizophrenic equivalents" and this term sums up the theoretical position adopted by many contributors to that volume. The fact that, in monozygous twin-pairs discordant for "schizophrenia" the "schizophrenic" co-twin has often been passive and asocial, with a low birth weight and a poor school performance as compared with his "un-schizophrenic" co-twin sibling (Kringlen, Pollin et al, 1966; Tienari), is also taken as confirmation of the broad-spectrum concept. Above all, the now well-attested finding that the first-degree relatives of "schizophrenic" patients are characterized by a wide variety of conditions, including positively valuable traits such as artistic and creative ability as well as mental disorders of all kinds (Heston, 1966; Schulsinger, 1972), suggests that a narrow concept may be misleading if rigidly applied.

It is not within the scope of this report to consider the genetic basis of "schizophrenia", the merits of a single gene versus a polygenic theory (Slater, 1958), or the other theories of causation that have been proposed. Our concern here is with the current status of the concept of schizophrenia. Throughout the discussion above the term "schizophrenia" has been placed within inverted commas because it is rarely precisely clear what the various investigators have meant by it, even though such an understanding is fundamental to any interpretation of their results. This deficiency is, of course,

most obvious where the less typical varieties of "broad-spectrum" schizophrenia are concerned. Until careful and reliable investigations of these varieties of "schizophrenia" have been made, the status of the wider concept must remain uncertain. If borderline forms of schizophrenia cannot be identified with accuracy, it is not possible to determine exactly how they are distributed in the population or how they are related to the "central" forms. There is no intrinsic value in broadening a diagnostic concept.

The most concerted efforts to define a "borderline syndrome" are those of Grinker and coworkers (1968). Their series consisted of young in-patients without delusions, intellectual deficits, alcoholism, or drug abuse, who had been repeatedly admitted because of "florid attention-provoking histrionic episodes", and who showed "an ego-alien quality to any transient psychotic-like behaviour". A lengthy but unstandardized record was made of interviews between caseworkers and patients and of the patient's behaviour. This record was then rated by social workers in a standard way and the ratings were subjected to a cluster analysis resulting in four groups: "psychotic borderline", "core borderline", "adaptive, affectless, defended, 'as if' persons", and "neurotic borderline". Relatively little additional experimental work has followed which would provide a basis for investigating the kinds of issue with which we have been concerned in this section, but at least a beginning has been made. Until more definite results are forthcoming, there is likely to remain a cleavage of opinion between those who prefer to set strict limits, and those who prefer to set broad limits, to the concept of schizophrenia. Both positions are scientifically respectable in so far as they can be clearly stated and defined, and thus communicated to others.

It may be worthwhile, before concluding this discussion, to consider the view that "schizophrenia" is not a disease at all but that some other theoretical model should be used to deal with whatever problems individuals who are given these appellations, have in common. Such a consideration, however, requires that the disease theory should first be stated. Kraepelin defined a disease entity in terms of unity of cause, course, and outcome. This is the physical disease model, in which each illness is recognizable in terms of a pattern of symptoms occurring together or over a period of time. The clinical condition results from underlying physiological dysfunctions due, in turn, to imbalance in some naturally-occurring homeostatic cycle. A full knowledge of etiology would require an acceptable theory of the mechanisms that keep such functions within normal limits, of the factors that upset the normal balance, and of the ways in which pathology and symptom-formation result. If the condition remains untreated, there is likely to be a characteristic course and outcome. One outcome may be that additional diseases develop. Intervention, whether purely empirical or based upon a knowledge of the etiology and natural history, may prevent the disease from occurring, cure it, suppress its manifestations, or influence its course (Lewis, 1953; Hare & Wing, 1970).

Thus the disease theory, when properly used, provides the clinician with a powerful tool. There is little wonder that generations of psychiatrists have attempted to apply it to mental disorders. Whether the theory fits or not, in any particular case, is an empirical question. What is needed to support or reject the disease theory, as with any other theory,

is evidence that has been carefully collected according to well-defined rules. At the least, when a clinician makes a diagnosis he can test predictions concerning the etiology, the pattern of symptoms and signs, the pathology, and the response to specific therapies or the natural course without treatment. Thus he can confirm that the disease is present or demonstrate that it is absent. To make a diagnosis is in fact to put forward a theory that usefully predicts these factors and in his everyday practice the clinician is constantly testing such theories. When he makes a psychiatric diagnosis, he has fewer tests at his disposal. He cannot usually be very confident about the etiology or the pathology, though he can be somewhat more specific about treatment and can usually make a prognosis. Because of the paucity of tests available, the scientific basis for making a psychiatric diagnosis is often less certain than it is in other fields of medicine. However, the fundamental assumptions are the same and they must be tested, in so far as possible, against the same standards.

Since psychiatric diagnosis depends so heavily on symptomatology, and since the description of the patient's clinical condition depends so greatly upon the interview, it is evident that one of the major ways of facilitating scientific investigation is to devise a method whereby symptoms can be elicited, recorded, and classified in a systematic and reliable fashion. This became one of the main aims of the IPSS.

There are, of course, other medical models (that of rehabilitation, for example) that draw on different concepts (Wing and Brown, 1970) and that should not be confused with the disease model. The question of whether the latter can usefully be applied to schizophrenia, or to other mental diseases, could be answered by the research that would soon be undertaken once the fundamental goals of ensuring reliable and communicable description and classification were achieved.

Before examining earlier attempts to apply an operational definition of schizophrenic conditions, the literature on the prevalence and distribution of schizophrenia will briefly be reviewed.

2.3 Epidemiological Studies of Schizophrenia

The value of any study of the distribution of schizophrenia obviously depends on what criteria are used to define cases and on how thoroughly the population at risk is screened. The marked variability in the concept of schizophrenia as used by various investigators has already been discussed. Variability also arises, as will be seen in the next section, because of a lack of standardization of interviewing techniques. Some clinicians cross-examine, others do not; some are content when a hypothetical diagnosis appears to be confirmed, others explore all possibilities thoroughly; some have full information from relatives and friends and clinical investigations available, others have to make do with a brief interview with the patient. Yet another difficulty arises when only those patients who have been referred to a particular psychiatric service are included in the study. Most of the literature concerning international comparisons of the psychoses is based on work in which neither interviewing nor classifying procedures were standardized or reproducible, and in which psychiatric services were not uniformly

available. The results of such comparisons must therefore be interpreted with caution. Nevertheless, there have been some interesting studies of various degrees of sophistication that provide a foundation on which to build a more scientific epidemiology. Two examples will be briefly mentioned.

The first is a survey made by Eaton and Weil (1955) of an Anabaptist sect, the Hutterites, living in small closely-knit farming communities in North America. Their religious traditions dated from the sixteenth century and had been cemented by prolonged persecution and consequent migration. Property was owned in common, and everyday life was simple, austere, well regulated, and pious. Families were large since there was no birth control, but there was no poverty either, and practically no crime or violence.

It was thought by some that such rural peace, community support, hard work, freedom from urban stresses, and good order would provide conditions in which mental illnesses would be most unlikely to develop. Eaton and Weil surveyed the various colonies — a few intensively, the others more briefly -- and concluded that about 6 per 1000 of the total population of 8500 had at some time suffered from a psychosis. This figure was not very much lower than that arrived at in other surveys in Europe and North America, except that most of the cases were due to depressive psychosis, while schizophrenia was relatively rare.

Clearly, the methods of case finding in this survey were subject to the clinical idiosyncrasies mentioned earlier, and the screening of the population was not particularly thorough, so the results cannot be taken at face value. In addition, the movement of persons out of the Hutterite colonies over the previous 20 years or so could not be determined with any accuracy. Murphy (1968b) has "calculated the 1961 schizophrenia admission rate for the Mennonites and Hutterites combined, from Canada's prairie provinces, and that rate is not significantly below average." Eaton and Weil's survey showed, as others have done since, that communities believed to be relatively free from psychosis because their conditions of life seem ideal to some observers do not live up to expectations when a more formal examination is made.

The second example concerns a quite different area — the far north of Sweden, where the climate is severe, summer is very short and for six weeks during the winter the sun does not rise at all. The population lives by small-scale farming and lumbering, communications with the rest of Sweden are poor, and many families live under very primitive conditions. Böök (1953), who carried out this survey, used a variety of sources to find cases, including mental hospital admissions, parish registers, the records of district physicians, and information from key people in the settlements concerned, such as parish clergymen and schoolteachers. Böök made the diagnosis in each case himself according to certain principles, of which he gives a general account.

According to his calculations, the incidence of functional psychoses was some three times higher than among the Hutterites but, more striking still, schizophrenia accounted for 85% of the cases while manic-depressive psychosis was almost non-existent. Böök's explanation is in terms of

genetics and selective migration. He feels that a schizoid personality is an advantage for survival in such areas, while people with a high risk of developing manic-depressive disorders are likely to emigrate.

These two surveys, with their completely opposite findings, are typical of the best of the early work, in which the case finding instrument was ultimately a single clinician. Neither has been repeated as yet.

Other surveys were carried out by teams of workers using an agreed upon method. For example, following Lin's epidemiological survey of Chinese communities in Taiwan (1953), the same research team conducted a comparative survey of the prevalence and clinical characteristics of various mental disorders in Formosan aborigines and Chinese (Lin and Lin, 1962). Schizophrenia was reported to be significantly lower among the aborigines than the Chinese and was characterized in the former by acute excitement, hallucinations, and confusion. The clinical course was marked by an acute onset, short duration, and an apparently benign prognosis. This study of aborigines was somewhat limited in significance because of the fact that the field survey was conducted in a post-war period when the population was still suffering from serious health hazards and when many of the psychotic patients may have died of infection, disease, and malnutrition and thus have escaped detection. A. Leighton et al. (1963) in Nigeria used a methodology similar to that of the Stirling County study (Leighton, D.C. et al., 1963) and provided comparable information for cross-cultural comparison. Schizophrenia did not receive major attention because of the small numbers involved.

The survey by Hagnell (1966) in Sweden provides further evidence, in addition to that of A. Leighton and co-workers, that rural communities have their full share of psychiatric disorders. Indeed, there is no evidence beyond the anecdotal that any society, whatever its living conditions, is free from mental illness. A good review is provided by Rueck and Porter (1965).

Much other work has been based on the statistics of people consulting psychiatrists. Murphy (1965, 1968a, 1968b), for example, has summarized the evidence that the incidence rate of schizophrenia is relatively high in the Tamils of southern India, the southern Irish, and the north-west Croatians, and has taken the analysis further in his studies of Canadian mental hospital admissions. He finds that Catholics appear to be particularly susceptible but that culture has an evocative rather than a simply distributive effect. His studies are particularly interesting because of his attempt to link epidemiological with laboratory work, an attempt that has parallels with the theories of the MRC group (Venables, 1968; Venables and Wing, 1962; Wing and Brown, 1970).

Ødegaard (1952) has argued that most patients with schizophrenia are likely to be hospitalized at some time in their lives and that, if allowance is made for the delay elapsing between the onset of illness and first admission, estimates of incidence and prevalence based on hospital statistics are similar to those based on population surveys. Certainly, similar rates of schizophrenia have been reported from all over the world by investigators using both types of method (Dunham, 1965; Fremming, 1951; Strömberg, 1938).

There are, however, some difficulties. For example, in a comparison between case registers established in Aberdeen, Scotland, in Camberwell,

London, and in Baltimore, Maryland, USA, based on the unduplicated statistics of in-patients, day-patients, and out-patients in these places, schizophrenia was found to be more prevalent in the American than in the British localities (in both sexes, and particularly between the ages of 25 and 64), while depressions were commoner in Britain at all ages, but particularly in women (Wing et al., 1967; Wing and Bransby, 1970).

This difference between the rates in the USA and in the UK, which had been noted several times before, was documented thoroughly by Kramer (1963, 1969b) using age-adjusted first admission rates to mental hospitals in the two countries. The American rate of first admission for schizophrenia to public and private mental hospitals in 1960 was 24.7 per 100 000 total population, compared with 17.4 for England and Wales. The rates for the major affective psychoses, on the other hand, were 11.0 and 38.5 respectively. Data for reactive depression were not available in the American figures.

Earlier studies in the USA, for example, by Lemkau and co-workers (1941) in Baltimore, Hollingshead and Redlich (1958) in New Haven, Connecticut, and Malzberg (1940) in New York, had not found such an excess of schizophrenia as compared with other parts of the world. The possibility of a real difference, however, appeared to be worth investigating.

Other social factors that have been found to be closely linked to the prevalence (and, to a lesser extent, the incidence) of schizophrenia have been geographical mobility and isolation (Dunham, 1965; Hare, 1956; Stein, 1957), occupational and social class (Kohn, 1968; Norris, 1956; Ødegaard, 1956; Murphy 1965; Goldberg and Morrison, 1963), and marital status (Brooke, 1967; Ødegaard, 1946; Norris, 1956; Malzberg, 1940). The significance of all these factors will not be reviewed here except to point out that their relationship to the social course, prevalence, and distribution of schizophrenia is generally accepted, while their relationship to cause and incidence is still disputed. The difficulty of using such concepts meaningfully in international studies need hardly be emphasized, particularly when complex indices such as social class are used.

Another area for investigation is the effect of social factors operating within the family on schizophrenia. There is now a substantial body of quantitative work indicating a relationship between such factors and the onset, course, and outcome of schizophrenic disorders (Brown et al., 1962; Brown et al., 1972; Wynne, 1968), although the question of what is cause and what is effect remains obscure, and the usual problems of diagnostic interpretation are manifest. There is also good evidence to show that a variety of social factors commonly help to precipitate onset and relapse in such disorders and are of great importance in determining their course (Brown and Birley, 1968).

In addition to such social factors, other obvious sources of variation in rates are differences in age or sex distribution, the genetic composition of the population, the pattern of occurrence of physical diseases or malnutrition, the prevalence of drug or alcohol abuse, and so on. Any survey that relies on morbidity statistics for case-finding must also take into account the innumerable selective factors at work as well as the differences in the quality and quantity of provision of services (Kramer, 1969b; Rawnsley et al., 1972; Wing et al., 1972). In many countries, the prevalence rate of

schizophrenia receives a substantial contribution from long-stay institutionalized patients, and the rates reflect both the fact that such institutions exist and that they influence the course of the condition (Gruenberg, 1966; Wing and Brown, 1970).

The incidence, course, and prevalence of schizophrenia are therefore likely to vary in different parts of the world simply because of the operation of biological, social, and administrative factors of these types, which are differentially distributed in various populations. Murphy's view that cultural differences operate through other social factors has already been mentioned. He has suggested, on the basis of a study with Raman in Mauritius (Murphy and Raman, 1971), that as compared with a British series (Brown et al., 1966) the course of schizophrenic psychoses in Mauritius is less chronic. A similar difference was suggested by the British authors, who compared the prognosis of their own series with that of other European groups described earlier, and it is possible that many of the factors concerned are amenable to social intervention. The second volume of this report will be more concerned with such matters than this one.

Lastly, it is inevitable in any discussion of differing rates that the subject of case-finding should again be taken up. The two major epidemiological problems are the reliable identification of cases and the satisfactory sampling of the population at risk; the former being concerned with the accuracy of the numerator and the latter with the accuracy of the denominator in any computation of rates. The next section will deal with factors leading to variability in making a diagnosis, but it is relevant to consider here the pathoplastic features that are likely to be particularly important in transcultural studies. For example, Dembowitz (1945) described a condition of acute excitement in Africans characterized by "restless confusion, violence and wandering". Gelfland (1957) described transient maniacal attacks; Carothers (1953) "frenzied anxiety"; and Field (1960) "fear psychosis". Lambo (1960, 1965), discussed the differential diagnosis of acute excitement. Similar conditions have been described in other parts of the world. The scope of variability in diagnosis, particularly if the condition has to be fitted into the International Classification of Diseases, is obvious. A case can be made that such conditions should be diagnosed as mania, schizophrenia, anxiety, or hysteria, as well as "Reactive Excitement" (ICD). The Windigo psychosis of the Chippewa, Ojibwa, and Cree Indians illustrates the way in which such states can develop. Lewis (1953) points out:

"These people suffer terrible hardships in the severe winters of north-east Canada. The scarcity of game obliges each family to live by itself, exposed to the risk of starvation; cannibalism sometimes occurs. They have myths about a monster, living as an ice skeleton during winter and dying in the spring, who devours human beings. They believe also that human beings may be led, by witchcraft, to develop similar cannibal desires and to have their heart turn to ice. In fact, some members of the tribe do become profoundly depressed and excessively anxious about starvation. Their perceptions then become disturbed and they see members of their family as plump, succulent, inviting beavers. Some of those affected have insight into their condition and beg that they should be

killed before they give way to their cannibal urge; others actually kill and eat members of their family and eventually other people if they are not caught in time".

Yap (1967) considers that "culture-bound reactive syndromes" are fairly common in some countries and can be classified within a conventional nosological scheme by psychiatrists who are sufficiently familiar with local conditions. A further point is that psychiatrists may develop a bias towards diagnosing conditions that are particularly common in their own culture, even when these cultural conditions do not obtain. This would be particularly important in a study such as the IPSS. In spite of the apparent similarity of many of the crude prevalence rates of schizophrenia quoted for different parts of the world, there is a possibility that these rates are made up differently in different areas. Detailed attention to precise measurement of the relevant factors in both numerator and denominator may well yield a rich harvest of scientific results. Katz et al. (1969b) have shown that, even where diagnosis is agreed upon, patients from different cultural backgrounds can present with widely different symptom pictures. This could make it even more difficult to develop comparable diagnoses across cultures.

We shall be concerned in this volume mainly with the numerator of incidence and prevalence rates. It is therefore necessary, before concluding this brief survey of the literature, to give more detailed consideration to efforts at standardizing the processes of psychiatric diagnosis.

2.4 Operational Definition of Schizophrenia

Introduction

The factors that contribute to the variability in diagnosing schizophrenia may be divided into those factors affecting the patient, those affecting the clinician and his methods of measurement, and those affecting the situation in which the diagnosis is made. The characteristics of patients include their attitudes to and experience of mental illness and the persons who attempt to treat it, their ideas about the nature of treatment, their social status, their desire for help, and the nature and course of their illness. The characteristics of clinicians include their training, orientation, and skills, the ways in which they collect information, the kinds of data they consider important, the tests they consider relevant, and their rules of classification. The characteristics of the situation include the time allowed for examination, the circumstances under which it takes place (at home, in hospital, in public, under constraint), the attitudes towards mental illness and its treatment that are held by society, the local cultural subgroups, and the patient's relatives, the quality and quantity of services provided, the point in the course of the illness at which the patient contacts these services, and the specific relationship between doctor and patient.

Two aspects of making a diagnosis will be considered in more detail, since both point to a major component in variability that could perhaps be standardized. The first concerns the rules used by clinicians to classify

mental disorders; the second concerns the ways in which information about the patient's condition is collected.

The International Classification of Diseases

The Eighth Revision of the International Classification of Diseases (ICD) was published in 1967, after many years of work at national and international levels (WHO, 1967a). Section V of the ICD contains 26 main groups covering mental disorders and mental retardation and attempts to meet the criticisms that had been made of the corresponding section in the Seventh Revision (Stengel, 1967). Several countries have published their own glossaries of the terms used in the Eighth Revision in order to provide operational definitions (e.g. GRO., 1968, APA, 1968).

The introduction to the Manual for the Eighth Revision states that "the purpose of a statistical classification of diseases is principally to furnish quantitative data that will answer questions about groups of cases". The ICD still, however, suffers from certain basic deficiencies that hinder the achievement of this excellent goal. There is a lack of consistency in the principles of classification, since several different axes (symptomatic, etiological, social, prognostic, etc.) are used somewhat haphazardly. The categories used are neither mutually exclusive nor jointly exhaustive (Hempel, 1959).

A detailed discussion of the problems still remaining in the Eighth Revision, and also in the accompanying British Glossary, is given by L. Wing (1970). Her criticisms of the rubric Schizophrenia* (295), which is separated from Paranoid States (297), including Paranoia and Involutional Paranoia, and from Reactive Psychoses (298), including Reactive Confusion, (Reactive Excitation, and Acute Paranoid Reaction) are particularly relevant to the topic under consideration. In the British Glossary, group 295 is introduced with a fairly detailed description of the symptoms said to be characteristic of schizophrenic illnesses. These include the subjective experience of disordered thought leading to delusions of influence and reference and hallucinations, particularly of voices repeating the patient's thoughts or commenting on his thoughts and actions, all in a setting of clear consciousness. It is not made plain, however, whether these symptoms need to be present in order for an illness to be classified as schizophrenia. The subgroups include "Simple schizophrenia" (295.0) and "Latent schizophrenia" (295.5), and the notes on both of these imply (but do not specify) that these diagnoses can be made without any of the above "characteristic" symptoms being present at all, either at the time of examination or in the past. Similarly, the notes on the subgroup "Catatonic schizophrenia" (295.2) describe the psychomotor manifestations of catatonia but do not provide any guidance on how to classify those patients in whom the motor phenomena appear without evidence of thought disorder, delusions, or hallucinations. A further difficulty is that there is no instruction as to how to classify a patient who has schizophrenic symptoms but has also been drinking alcohol excessively or taking amphetamine. After several years of work WHO has now prepared an international glossary for use with the 8th Revision of the ICD.

Ward et al. (1962), studying diagnostic agreement on 153 cases between

* Numbers in parenthesis refer to categories of the ICD, 8th Revision, 1967.

four psychiatrists, reported that two-thirds of the discrepancies were due to inadequacies in classification. (The American Psychiatric Association DSM I was used). Approximately one third was due to inconstancy of the observers, and 5% due to inconstancy of the patients.

Kreitman (1961) showed that the degree of diagnostic agreement between psychiatrists varied very widely with the type of condition studied, the setting in which reliability was assessed, and the background of the psychiatrists. At best, there appeared to be good agreement on specific organic disorders such as general paresis (90%), epileptic psychosis (92%) and mental retardation (91%), a lesser degree of agreement on the functional psychoses (69%), and rather poor agreement on the neuroses (24%). At worst, there was barely any agreement at all. In his own study of the diagnoses of five psychiatrists, he found 85% agreement on the organic psychoses, 71% on the functional psychoses, and 52% on the neuroses, taken as broad classes (Kreitman et al., 1961).

Such results have been paralleled in studies of agreement on physical diseases. In a recent paper, Norden and coworkers (1970) described an experiment in which doctors were asked to examine a series of intravenous pyelograms and decide whether the patient had chronic glomerulonephritis: the inter-rater disagreement was 20 per cent and the intra-rater disagreement 10 per cent. These findings confirm those of earlier studies in which raters disagreed in reading chest x-ray films (Birkelo et al., 1947) and on whether the patient had finger clubbing, both in live observation and in photographs (Pyke, 1954). Perhaps the international studies of the reliability of rating symptoms such as cough and sputum, and of diagnosing chronic bronchitis, are the most relevant examples for our own concerns.

Problems in the definition of the "functional psychoses"

Probably one of the most important sources of diagnostic variability is the lack of agreement between clinicians on the rules of classification. Thus, the term "functional psychosis" has no precise meaning apart from specifying the absence of any obvious organic etiological factor. It is useful only as a label for a group of conditions characterized by disorders of thought, perception and affect, often elaborated into delusions or hallucinations, and accompanied by a wide range of behavioural disturbances. Symptoms such as clouding of consciousness, disorientation, and amnesia are typically absent. The diagnosis rests solely on the skill of the clinician in eliciting present and past symptomatology, interpreting any clues as to causation, and applying the rules of classification he has been taught in the school of psychiatry to which he belongs. Any international comparisons are thus beset with manifold difficulties. These may be summarized under five headings, as follows:

(i) Differentiation between psychotic and non-psychotic conditions.

In the first place, there is disagreement as to where to draw the line between psychotic and non-psychotic conditions. Indeed, there is even disagreement as to whether any line exists at all. A respectable case can be

made for the proposition that there is an endogenous or psychotic form of depression, distinguishable from other forms in symptomatology, outcome, etiology, and treatment. At least an equally strong case can be made for the opposite view, namely that psychotic and neurotic forms of depression lie on the same continuum (or continua), one shading imperceptibly into the other. Similarly, there are schools that diagnose simple or pseudo-neurotic forms of schizophrenia in the absence of the more obvious symptoms characterizing the hebephrenic, catatonic, or paranoid forms, while other schools would not make a diagnosis of schizophrenia in such cases. Closely allied to such differences in point of view are differences concerning the severity of symptomatology that can rightly be termed psychotic. Is mild hypomania, with some euphoria and flight of ideas but no delusions or gross behavioural disturbance, to be called psychotic or not? In practice, clinical decisions are bound to vary from one patient to the next.

(ii) Combinations of different forms of psychosis.

Secondly, even the generally accepted psychotic syndromes occur in combination as well as in relatively pure form. Thus, delusions of passivity, which are usually regarded as typically schizophrenic, may be accompanied by delusions of guilt, which would be regarded as typical of psychotic depression if they occurred in a different setting. There is no uniform procedure for deciding the diagnosis in such cases.

(iii) Influence of etiological factors on diagnosis.

Thirdly, there are variations in the way possible etiological factors are dealt with in arriving at a diagnosis (as well as variations in the thoroughness with which their presence or absence is investigated). For example, syndromes indistinguishable from schizophrenia, with no evidence of toxic symptoms, may sometimes follow amphetamine or alcohol abuse. Some clinicians would label such conditions "schizophrenia (drug-induced)", while others might prefer "alcohol (or amphetamine) psychosis" or other labels. Even when the supposed etiological factors are psychosocial in nature, as in a psychosis following bereavement, emigration or severe persecution, there are differences in opinion concerning how this affects classification. One school of thought predominant in Scandinavia assigns such cases to a special category of "psychogenic psychoses", but even within this school there are questions of how "reactive" a psychosis must be before it can be included under this rubric.

(iv) Present mental state and previous clinical history.

Next, differences in classification arise because of a lack of uniformity in dealing with the patient's previous experience of psychiatric illness. The difficulty is most obvious when there has been an earlier clear-cut episode of schizophrenia but where the patient presents on a subsequent occasion with some other condition that might not be called psychotic at all if taken by itself. Some clinicians would still diagnose schizo-

phrenia, others would not. This is one example of a general difficulty concerning the extent to which the present diagnosis should be determined by the earlier clinical history.

(v) Social and cultural factors influencing the diagnosis.

Finally, and particularly important in international comparisons, social and cultural influences on the clinical picture have an important effect on diagnosis. Clinical practice is rather variable in respect of these. For example, a West Indian man who had recently arrived in England, who belonged to a religious sect that practised some form of voodoo, and who had a poor command of English and perhaps a low intelligence would be likely, if he became excited or depressed, to present a clinical picture very different from that presented by a well-educated Englishman. The diagnosis would also be more likely to be schizophrenia. Such pathoplastic features were discussed in more detail in the previous section.

Differential perception of symptoms

It has been suggested that some differences in diagnosis are due to the fact that psychiatrists actually perceive different symptomatology when, for example, rating the same film or videotape. M. Katz et al., (1969a) found that flatness of affect in a filmed subject was rated present by American psychiatrists, who diagnosed schizophrenia, but rated absent in the same case by British psychiatrists, who diagnosed a neurotic or personality disorder. Kendell (1971) found that a videotaped patient was rated as having delusions, passivity feelings, and thought disorder by American doctors, who diagnosed schizophrenia, but not by British doctors, who diagnosed personality disorder. Another explanation of these results is that psychiatrists use different definitions of symptoms such as flatness of affect or delusions.

Non-diagnostic rules in classifying

Pokorny (1965) has listed several administrative or therapeutic factors that influence American psychiatrists to use a particular diagnostic label, even though these factors may be totally irrelevant to the true diagnosis. Thus, if the clinician thinks the patient should be committed to a mental hospital, the diagnosis will tend to be one of the psychoses. If the patient appears to need ECT, the diagnosis will tend to be one of depression. If the clinician decides on psychotherapy, he will tend to diagnose a neurosis. If the aim is to have an ineffective individual discharged from the army, the diagnosis will tend to be inadequate personality. Similarly, the diagnosis may be fitted appropriately to the course of action desired when a decision is being made concerning criminal responsibility, eligibility for disability compensation, or admission to hospital or another institution. Babigian et al. (1965) showed that when a patient is moved from private practice to other types of medical service, this is often accompanied by a change of diagnosis from various non-psychotic diagnostic groups to a psychotic one. He also established that the diagnosis varies with the same rater and patient,

according to the purpose for which the patient is seen (e.g., diagnosis or treatment).

To the extent that such practices are widespread, they vitiate the usefulness of diagnoses. More study is needed of these factors to determine the extent of their impact. If it is considerable there will have to be more emphasis on training psychiatrists in methods of diagnosis if epidemiological studies based on these diagnoses are to be of value.

2.5 Efforts at Clarifying the Sources of Diagnostic Unreliability

Recently, there have been several attempts to clarify the sources of diagnostic unreliability, for example, the WHO Programme A diagnostic exercise on schizophrenia in the London Seminar in 1965, and the US-UK diagnostic study.

WHO diagnostic exercise on schizophrenia

The first of the diagnostic exercises carried out as part of WHO's Programme A was concerned with schizophrenia (Shepherd et al., 1968). Twenty psychiatrists considered six written case histories and made a full diagnostic formulation as well as a diagnosis in terms of the ICD. The presenting illness appeared to be the most important factor on which the formulation of diagnosis was based, supported by details of past mental illness and assessment of personality. When these conformed to a characteristic clinical picture, the level of agreement was high. Where there were unusual clinical features, especially if these occurred in association with a physical disorder that might have been etiologically important, there was much more disagreement.

The psychiatrists also considered nine videotaped psychiatric interviews, in which the fifth edition of the standardized Present State Examination was used. The authors summarized the reasons for diagnostic disagreement as follows: firstly, variations in observation of symptoms and signs; secondly, variations in inferences from observations; and thirdly, variations in diagnostic terminology. However, there was a surprising degree of agreement on the presence or absence of specific symptoms, no doubt owing to the use of a systematic and precise interviewing technique. Unstructured assessments of affect produced most disagreement, as would be expected.

US-UK diagnostic study

Following the observations made by Kramer (1963, 1969b) on the discordant first admission rates for depression and schizophrenia into mental hospitals in England and Wales and the United States of America (see Section 2.2), a study known as the US-UK Diagnostic Project was undertaken to investigate the reasons for these differences.

Samples of newly admitted patients were interviewed in New York and London and a clinical history was taken. The interviewers made psychiatric diagnoses in terms of the eighth revision of the ICD, using the British

Glossary. They were not aware of the diagnoses made by the responsible hospital clinicians, which were subsequently obtained from the records of the British Ministry of Health and the New York State Department of Mental Hygiene.

The results of this exercise were fairly clear-cut. Hospital psychiatrists in New York did, as expected, diagnose schizophrenia more frequently, and affective disorders less frequently, than hospital psychiatrists in London. However, the psychiatrists of the project team, all British trained, diagnosed these disorders in much the same proportions in each city, and these proportions were very similar to those of the London Hospital psychiatrists (Cooper et al., 1972; Gurland et al., 1969; Kendell et al., 1971).

Subsequent videotape studies substantially confirmed these findings, which indicate that the prevailing concept of schizophrenia is much broader in the United States than in Britain, embracing substantial parts of what British psychiatrists would regard as depressive illness, neurotic illness, or personality disorder and almost the whole of what they would regard as mania (Kendell, 1971). Thus a considerable amount of literature on both sides of the Atlantic concerned with epidemiology, genetics, family psychopathology, drug treatment, and rehabilitation needs to be reconsidered, since British and American results cannot be assumed to be comparable.

This study demonstrates that comparing patients from different centres on the basis of local diagnoses alone can be very misleading and that standardization of data collection and diagnostic processes is required.

2.6 The Development of Standardized Data Collection Techniques

Early attempts at cross-cultural studies of psychopathology had to depend for collection of data on correspondence with local psychiatrists (Wittkower, et al., 1960) or the shifting of a research team from one centre to another (Leighton, A. et al., 1963; Leighton, D.C. et al., 1963). Realizing some of the limitations of these pioneering efforts, investigators have more recently turned to the use of standard schedules for data recording to assist in collection of comparable data in different centres. Lorr and Wittenborn have developed instruments for rating a wide range of psychiatric symptomatology and have demonstrated their reliability. A considerable number of rating scales have been developed and used with variable success. The scales differ in their aims, length, structure, scope, and other characteristics (Lyerly and Abbott, 1966; Scharfetter, 1971).

More recently, attempts have been made to increase data comparability further by standardizing the interview procedures for eliciting data from the patient and informants as well as the procedures for recording such data. Basically, a psychiatrist makes use of three kinds of information when making a diagnosis: (1) A "present state" examination, which may be repeated several times, in order to discover the morbid phenomena experienced by the patient and any evidence of abnormality in his behaviour and mode of relating. (2) A clinical history, taken from the patient and other sources, which indicates what symptoms have occurred during the present and previous episodes of illness. (3) Other historical information, and pathological and

other investigations that may have a bearing on the etiology of the condition. The discovery of amphetamine in the urine, a brain tumour, or a history of heavy alcohol intake may obviously have a decisive effect on the diagnosis. On the whole, however, psychiatric diagnosis is based upon the interview and any attempt at standardization must begin at this stage.

An attempt to standardize a psychiatric interview has been made by Spitzer and co-workers (1964, 1967, 1969). This technique specifies both the questions to be asked by the examiner and the way in which the answers are to be coded. Much use is made of nonspecific probes, such as "Can you tell me more about that?", to draw out material from the subject, but there is no cross-examination on phenomenology. Many of the symptoms are, in effect, rated by the patient rather than the examiner. Reliability has not been extensively investigated but appears to be acceptable. Analysis has followed both dimensional and diagnostic lines. Of course, thinking that determined the selection of items and the conduct of the interview was American rather than European. No results of international studies have yet been published, nor has the technique been used specifically to investigate the processes of clinical diagnosis.

Another technique of this kind was published by Wing and co-workers (1967, 1970, 1971) after five years of development in the Medical Research Council Social Psychiatry Unit. It consists of a partially standardized interview, the Present State Examination (PSE) — which at that time had been through five editions —, and a syndrome check list based on the PSE that can be used to rate previous clinical episodes. This procedure employs the common model of a psychiatric diagnostic interview based upon cross-examination.

The clinician has in mind a particular symptom and, in order to be certain whether the patient has it or not, he will ask a series of questions, each one depending upon the patient's previous reply. This is a flexible and efficient procedure that is precluded by the use of a questionnaire. The principles of the interview are described in more detail in Chapter 5, and the reliability of the procedure is examined in Chapter 8 (see also Wing et al., 1967; Kendell et al., 1968; Sartorius et al., 1970). As mentioned above, the PSE was used by the US-UK Diagnostic Project (Kendell et al., 1971).

These methods show that, for research purposes at least, interview variability can be reduced by standardizing the clinician's interviewing and other data-gathering techniques. The next step then is to standardize the classifying rules and their application to the data collected.

2.7 The Development of Methods to Translate Raw Data into Psychiatric Diagnoses

Three methods have been used to translate raw data into diagnoses; these attempt to solve the problems of both diagnostic criteria and diagnostic reliability.

(1) Clinical method. In this method the psychiatrist merely uses the collected data as a source of information for arriving at a diagnosis

according to his own concepts. Such a method probably is not highly comparable among psychiatrists, especially if they are from different centres.

(2) Computer diagnostic programs. Such programs (e.g., Spitzer's DIAGNO, Wing's CATEGO) use raw data as inputs to a program written to make successive decisions as to the diagnostic category appropriate for the patient. The principles for decision-making are based on the psychiatric concepts of the psychiatrist who develops the program, but they are applied, through use of the computer, in identical ways to all the cases analysed.

(3) Factor analysis and clustering techniques. The need for a fresh approach to the classification of psychiatric disorders was early apparent to psychiatrists, perhaps particularly to those working in the USA. Lorr and co-workers (1963) put forward a strong case concerning the need for such a new approach. As Lorr (1966) put it, "in much of American psychiatry, formal diagnosis is actually ignored as relatively unimportant and outmoded, or disparaged as nondynamic and useless". The main reasons for dissatisfaction were disagreement between diagnosticians, the high proportion of unclassifiable cases, and the lack of validity. The latter stricture applies, of course, even more strongly to any new system that might be devised. Lorr applied factor analytic techniques to ratings of patients' symptoms. This provided new groupings that had thus been developed independently of any particular disease theory. These also had the advantage, as with the computer methods discussed above, of being able to assign new patients to diagnostic groups with absolute reliability. Clustering methods of classification provide yet another way of defining patient groups based on mathematical principles. These can also provide perfectly reliable assignment of patients to groups. They will be discussed further in Chapter 12.

MANAGEMENT AND OPERATION3.1 Management3.1.1 Introduction

When beginning a research project, particularly of the magnitude and complexity of the IPSS, it is usual to seek out reports of similar projects so as to learn from the experience of other investigators. Unfortunately, the literature on transcultural research contains very little about the practical aspects of management.

This lack of concern, whether real or apparent, with management problems is reflected in the fact that it is rare to find a group of scientists who, when planning a study, invite a management expert to advise them on what sorts of difficulty they will meet and how best to overcome them. There are still very few studies in psychiatry or other branches of medicine in which special arrangements are made for the routine recording of the management and operational aspects of the study while it is being carried out. If any description of management procedures is given, it is made retrospectively and somewhat cursorily so as not to bore the reader before he gets to the research findings. Although there are some "cookbooks" in the field they do not include the results of experimental testing of management procedures, but either specify how the investigator thinks the project should be managed or else relate what actually happened.

This unfortunate lack of information means that in new projects management schemes have to be experimental or the work has to proceed by trial and error, which may increase the cost of a project. Thus, when the IPSS was begun little was specified about the project's management aspects in the early documents and descriptions of work. As a result, management had to be experimental, elastic, and capable of adjustment to adapt to all the possible obstacles and disturbances that might arise.

If the chief investigator can carry out his project by travelling from one research centre to another, his continuing attention may compensate for the lack of a pre-determined management plan. While this method has many advantages, it is not applicable to research studies in which the cultures involved have different languages, where the distances are very great, when the number of patients exceeds a minimum, and where investigations are to be made simultaneously in a number of countries. The IPSS had all of these characteristics, and producing an adequate management strategy was as great an objective for research as the actual topic of the study.

3.1.2 Administrative structure and considerations

In addition to providing means of resolving difficulties that could be expected to occur in any project involving the detailed examination of a

large number of patients, the administrative structure of the IPSS had to be equipped to solve problems arising from several other sources.

Some problems arose from the international character of the study. Inter-governmental relationships are not always favourable for research and may, on occasion, thwart the carrying out of activities essential for the conduct of a study. If some of the countries taking part do not have normal diplomatic relations, there may be so much delay in obtaining visas and other documents that collaborating investigators cannot attend meetings. Strict import or export regulations may slow or prevent the movement of research equipment. While some regulations are known in advance and provisions can be made for them, for example, by special arrangements with governments, others crop up unexpectedly and cause even greater loss of time. Similarly, national statistical requirements, which vary from country to country, may have a significant influence on the context and organization of hospital records and hence on the availability of information about the patient and the content of the research schedules. National legal provisions, e.g., for studies involving human subjects, may similarly affect the research procedures.

The distance from Headquarters and from each other of most of the field research centres (FRCs) taking part in the IPSS created problems of communication, which were often aggravated by delays in postal services. The effect of distance was also felt more strongly because the administrative scheme purposely involved the Centres as much as possible in decision making. It can be said that the time needed to carry out a project activity is directly proportional to the distances between the centres in which the work is carried out.

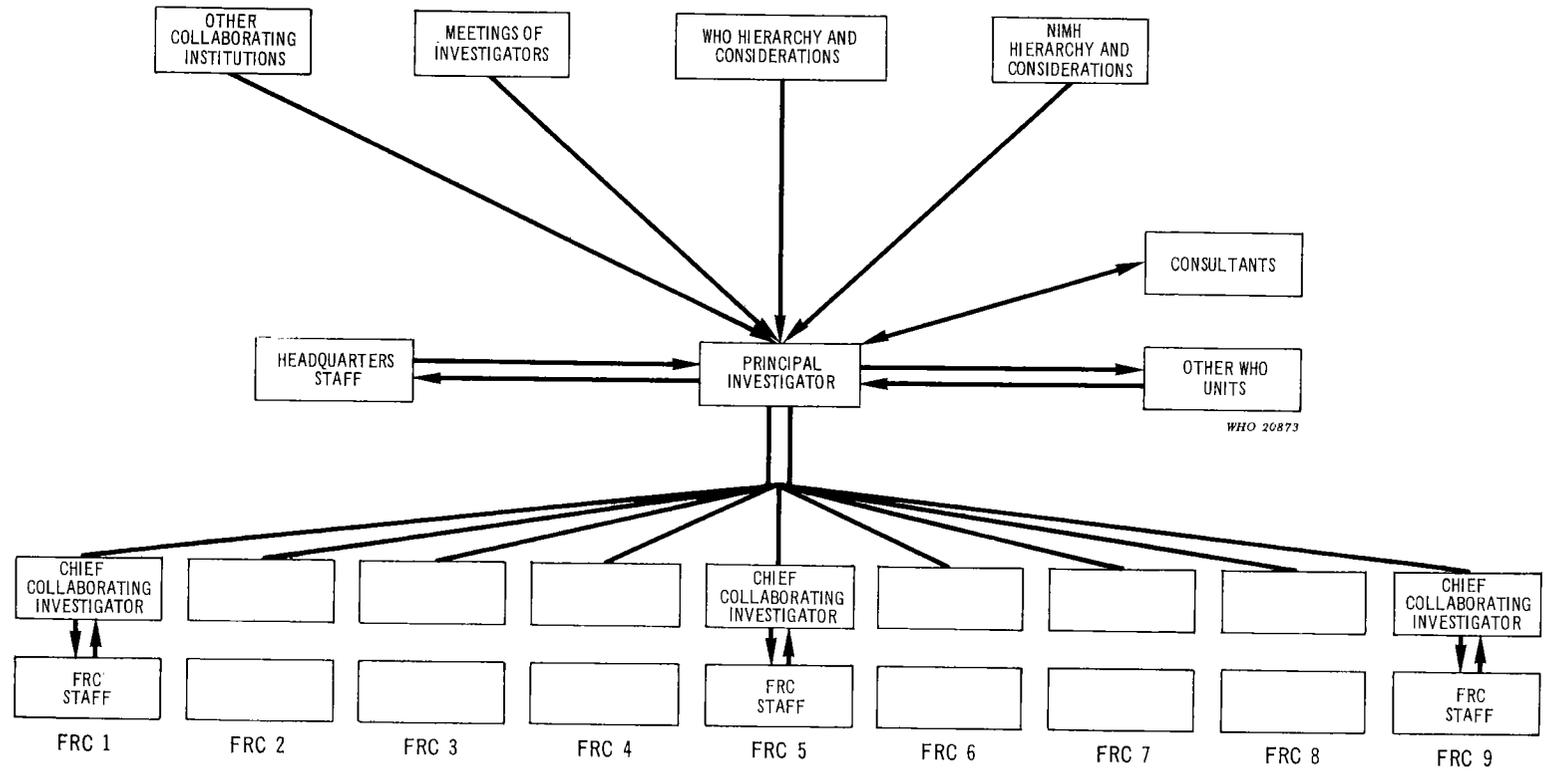
There were further administrative considerations stemming from the nature and aims of the two agencies that were sponsoring the project. WHO's carefully organized system for giving priority to the interests of member states must impose checks and controls on all its activities and compliance with these can result in administrative delays. The National Institute of Mental Health also has its own rules and regulations, and time has to be taken to ensure that these are observed.

The administrative structure of the project is shown schematically in Figure 3.1.

As regards the FRCs, each one had a chief collaborating investigator who in several cases was the most experienced psychiatrist in the Centre. Staffing plans also provided for one or more collaborating psychiatrists, a clinical psychologist, a social worker, a clerk, and supporting staff. In practice, the number of staff engaged in the IPSS varied considerably from centre to centre. In one there were only two psychiatrists and a secretary, whereas in another there were three social workers, two psychologists, a statistician, and three psychiatrists, who were all actively working on the project. Such variations were due partly to differences in working conditions in the catchment area (including the number of facilities from which patients were taken into the study), partly to the lack of trained personnel in some centres, and partly to the varying amount of research experience among the existing staff.

Although there were several centres that retained the same psychiatrists

FIG. 3.1 ADMINISTRATIVE STRUCTURE OF IPSS



throughout the course of the study, in others the whole psychiatric staff changed. In FRCs with staff turnover, the difficulties of training new staff and of checking and maintaining reliability while carrying on the routine interviewing, were compounded by the fact that investigators who have been trained together and reach a high degree of reliability in their ratings tend to deviate from this agreement when working alone, with only infrequent opportunities for checking on their reliability. When these investigators train other staff, the divergences are perpetuated. Similarly, when the newly trained investigators begin to work on their own the whole process is repeated, so that with time the divergence from methods established by the original training may become very great.

At Headquarters there were many changes of both professional and secretarial staff. This required time-consuming in-service training and sometimes cut off promising avenues of research.

Supporting structures

The structures supportive of this project can be discussed in only a general way because it is difficult to enumerate them and even more difficult to describe their roles in detail. The inter-governmental nature of WHO contributed most significantly to the feasibility of the study. It is hard to imagine that any other national or inter-governmental agency could have been a better vehicle for the project or that a private investigator could have obtained international collaboration on anything like the same scale. Many of the obstacles posed by national or political limitations were overcome thanks to WHO's unique position in the field of health care. However, as this advantage rests on an unblemished record of good relationships with member countries, it means that great care has to be given to anything issued officially from the Organization. This is made possible by meticulous control of the activities of the technical units in so far as their possible non-technical repercussions are concerned. The need for such checking demands time and imposes limitations that are often more easily condemned than understood and allowed for in planning.

The Headquarters of the IPSS was housed in WHO's Headquarters in Geneva and this ensured easy access to other parts of the Organization, such as Budget and Finance, Mailing and Travel, which were most helpful in the administration of the project. In addition, an effective collaboration with technical units, such as Data Processing and Health Statistical Methodology, furnished most valuable technical support. Similarly, specialists or structures in the FRCs provided essential support for the day-to-day operation of the project.

Typical of the use of supporting structures in the Centres were the simultaneous trials of various methods of computer grouping of patients carried out in three FRCs with computer facilities.

Communications

Because of the distances between the Centres and the complexity and timing of the project, it was of crucial importance to maintain communication

among the centres and with Headquarters. Although the following observations may seem trivial to some readers, the size of the resulting workload is not always appreciated; if proper provision is not made for secretarial staff, both skills and money will be wasted as professional staff spend their time on these activities in order to keep the project moving. The communication patterns in the IPSS fell under four headings:

- (a) correspondence; sending material and equipment;
- (b) visits of Headquarters staff to the FRCs;
- (c) exchanges of visits of collaborating investigators;
- (d) general meetings of investigators.

(a) Correspondence: sending material and equipment. Apart from special correspondence in connexion with the organization of (b) to (d) above, routine communication included the following:

(i) Circulation of draft schedules; collection of comments and suggestions from FRCs on drafts of schedules or research proposals produced at Headquarters.

(ii) Flow of records and other data from FRCs to Headquarters for processing and analysis, involving some 9,000 schedules.

(iii) Replies to queries connected with records; feedback of results from Headquarters to FRCs; exchange of letters in connexion with special studies carried out in the Centres.

(iv) Correspondence concerning supplies, equipment, apparatus, etc.

Delays of correspondence due to distance, technical difficulties and the busy time schedules of the collaborating investigators resulted in delays in all attempts to carry out changes of procedure, collect additional information, or make arrangements for meetings. They effectively prevented the circulation of films of psychiatric interviews for rating in the Centres. Had this been feasible, it would have been the best way of testing inter-rater reliability since all collaborating investigators could have taken part.

(b) Visits of Headquarters staff to FRCs. This method of communication would have been one of the best ways of promoting efficiency and uniformity in carrying out the study. However, although some of the Headquarters staff visited the FRCs from time to time, the long distances involved and the shortage of time at their disposal made these visits very brief. Some of the Centres, for one reason or another, were visited only once or twice in the course of the project. The purposes of the visits to the FRCs evolved in the course of the project. In 1966 and partly in 1967, the visits were designed to acquaint Headquarters staff with the Centres and their staff. In the following two years they were made to coordinate the functioning of the pro-

ject and to clarify matters that required on-site discussion.

Shortage of time prevented more extensive use of Headquarters staff in training members of the research teams in the Centres or in suggesting pertinent subsidiary studies that might have been developed locally. Headquarters staff could have acted as a group of "standard" raters whose rating method would have served as a model and whose ratings would have provided a yardstick against which the ratings of all collaborators could be compared. Another possibility would have been to give assistance in data analysis carried out locally, especially in Centres where no statistician was available before the project began.

(c) Exchanges of visits of collaborating investigators. These were held on several occasions and involved groups of varying sizes meeting in the Centres or at Headquarters. They fulfilled the following purposes:

- (i) enabled the collaborating investigators to review the activities carried out in the FRCs and Headquarters and to discuss operational and technical problems encountered in their work;
- (ii) provided experience in the use of the research instruments;
- (iii) acquainted the investigators with the methods and results of data analysis;
- (iv) enabled research plans to be formulated for additional studies designed to elucidate problems or test hypotheses based on data already available;
- (v) provided means for testing or improving inter-rater reliability within and between the FRCs.

During the first year of the project, most of the investigators had the opportunity to visit other centres. Those from Cali visited London, Moscow, and Washington; those from Agra visited Aarhus and London; those from Moscow went to Aarhus, Cali, and London; and those from Taipei visited Cali and Washington.

In the second phase of the IPSS, exchanges of visits grew into larger gatherings because, in addition to the purposes listed above, the investigators had to act as a consultative body to discuss changes in the research design as necessary and introduce new schedules or changes in the research procedure.

One of the most important benefits derived from the exchanges of visits was that they contributed to the development of a team spirit among the investigators and led to spontaneous decisions to collaborate among themselves in subsidiary studies. Important decisions about changes of research procedure could be discussed and reached without loss of time. It was possible to test and enhance inter-centre reliability during these meetings.

However, for practical reasons, it was not possible for the exchanges of visits to fulfill all of their potential. It was originally hoped that

all members of the FRC teams could participate in such exchanges, but because of the distances, number of people, language problems, and expenses involved, it was usually possible for only the psychiatrists on the teams to attend them. Although in each of the FRCs there was considerable discussion between the psychiatrists and other FRC team members, such as psychiatric social workers and psychologists, it would have been a great advantage if it had been possible for the non-psychiatrist team members of the FRCs to enter into face-to-face discussions with members of the other FRC teams. Not only would this have enriched the discussions at the exchanges of visits, but it also would have provided fuller opportunity for the research education of entire FRC teams and would undoubtedly have further increased the morale of the teams. Whenever possible, other team members were invited to such exchanges, and in addition the full staff of a particular FRC would be present when an exchange of visits was held at that Centre.

(d) Meetings of investigators. At these meetings, basic policy decisions were made, and therefore it was essential that every Centre be represented. They were most crucial at the beginning of the project, when the strategy of the study was agreed upon. One meeting held at Headquarters in Geneva in 1966 focused particularly on the aims, design, and methodology of the study and at another in 1967 the results of Phase 1 were presented and discussed in detail, and research instruments and procedures were finalized for use during Phase 2.

3.2 General Organization of the Study in the Centres

The following organizational procedure was generally accepted, with small modifications being made to suit local conditions, such as the characteristics of the catchment areas and the size and organization of the workload. The tasks of the various members of the FRC research team and the time required to carry them out are shown on the flow chart in Table 3.1.

All patients who contacted the selected psychiatric facilities were first seen by a clerk in charge of the administrative arrangements for the study. His task was to fill in the first (demographic) part of the Screen Form and to keep records of patients who were ineligible for the study because they lived outside the catchment area or were outside the age limits of 15-44 years, and also of those patients who passed the demographic screen. For the latter patients, the clerk gave the Screen Form to the first psychiatrist or medical officer to examine the patient so that he had the form when interviewing the patient in the ward or outpatient department. This psychiatrist filled in the second part of the Screen Form, indicating whether or not the patient should be included in the study. In one of the FRCs where there were a great many patients to examine, the clerk carried out a kind of pre-screening by reading the case-notes and referring to the psychiatrist those patients who might be eligible for the study. When the Screen Form was completed, it was returned to the admission clerk who made monthly reports to Headquarters on the numbers of patients screened and the reasons for rejection or inclusion.

For patients who were to be included in the study, one person, either the clerk, another member of the research team, or the principal collabor-

TABLE 3.1 FLOW CHART: TASKS OF RESEARCH TEAMS IN THE PROCESS OF ASSESSING THE PATIENT

Estimate of time spent per patient			
30'	20'	120'*	150'*
Clerk	Medical Officer	Collaborating Psychiatrist	Research Team Members
<p>Tasks:**</p> <p>Keep register of patients</p> <p>Apply demographic screen</p> <p>Ensure that screen 2 reaches Medical Officer</p> <p>Completion of monthly reports</p> <p>Checking and mailings of schedules to Headquarters</p> <p>Arrange appointments for out-patients to be interviewed in study</p>	<p>Tasks:</p> <p>Apply screen 2</p> <p>For some centres: fill in Physical and Neurological Examination Forms and Condition on Admission Form</p>	<p>Tasks:</p> <p>Carry out Present State Examination</p> <p>For some FRCs: Psychiatric History Interview Physical and Neurological Examination</p> <p>Contact with psychiatrist in charge and with informants</p> <p>Complete diagnostic assessment</p> <p>Optional:</p> <p>Special studies on subjects of particular interest</p>	<p>Tasks:</p> <p>Complete Psychiatric History Form and Social Description Form</p> <p>Home visits and interviews of informant to obtain heteranamnesis</p>

* More in reliability cases

** In some Centres done by other research team members (e.g., psychiatrists).

ating investigator, decided the order in which the various examinations should be performed. The order varied from case to case depending on the availability of the members of the research team, the condition of the patient, whether he was an inpatient or outpatient, etc.

The total assessment of a patient by all the instruments described in Chapter 5 took on the average about 5 hours, even longer if he were uncooperative. Reliability interviews, which were given to about one patient in six, also took more time.

The psychiatrist who was to conduct the examination with the Present State Examination (PSE) schedule was allowed access to any available information about the patient. If the patient was not cooperative or his clinical condition was such that no reply could be obtained to the parts of the PSE where reported symptomatology is recorded, the psychiatrist was entitled to interrupt the investigation and continue it later. At the exchange of visits in Cali in 1968, it was agreed that as many interviews as necessary should be done in order to obtain all the information required. Each new interview was to be done with a new schedule and all schedules were to be sent to Headquarters. Having completed the PSE, the psychiatrist could proceed to fill in the other schedules, provided the patient was not too tired or too disturbed.

The other schedules could be filled in by various members of the research team. For the Psychiatric History and Social Description schedules, an informant was to be interviewed in all cases where this was possible. A medical officer had to do the Physical and Neurological Examination and fill in the form. This was to be sent to Headquarters only if physical or neurological abnormalities were found.

After all the information had been collected and all the instruments completed, the psychiatrist was expected to discuss the case with the other medical members of the research team and then to record his diagnosis on the Diagnostic Assessment form.

All schedules and forms were returned to the clerk in the FRC who was responsible for sending them to Headquarters.

To maintain the reliability of assessment between observers in the Centre and to measure their reliability over time, simultaneous and consecutive reliability interviews were carried out. In simultaneous interviews, both interviewer and observer filled in PSE schedules and Diagnostic Assessment forms independently and then discussed their ratings together. These discussions of the ratings served to maintain the level of reliability reached in the training period. Simultaneous interviews were also used for in-centre training of psychiatrists who entered the project while the study population was being accumulated. All new members were required to do five simultaneous interviews with each of the project psychiatrists in the Centre and to discuss their ratings and diagnoses with them. Throughout Phase 2 the first interview in each month was a simultaneous interview, with the aim of maintaining a continuing check on the reliability of rating.

At the end of Phase 2, each FRC was asked to describe its catchment area, how the research plan had worked, the difficulties encountered, and the influence that their participation in the study had had on other colleagues working in the FRC facilities. Little mention was made by the FRC

staffs of difficulties encountered in carrying out the IPSS, although their reports were no doubt influenced in the usual way by the researchers' personalities, general attitudes, and degree of experience in research work.

In the Agra FRC, where the catchment area covered 52,076 sq km (20,342 sq miles), the work was seriously hampered by lack of means of communication, long distances, bad roads, and lack of money to pay the expenses of transport. Taipei also had some of these difficulties, although to a lesser extent. The Ibadan Centre differed from the others in that it was divided into two parts, Ibadan and Abeokuta, which are 96 km (60 miles) apart. The majority of the patients come from around Ibadan, where the administrative work of the FRC was carried out, but the larger inpatient facility consisting of Aro Hospital and Aro village is at Abeokuta. Although the Cali catchment area was relatively compact, 352 sq km (137.5 sq miles), many patients came into the city from the surrounding rural areas in search of medical treatment, and this increased difficulties of communication with them and their families. In Agra, Moscow, and Taipei, the feeling that mental illness was a stigma made patients and their relatives unwilling to come to hospital. In Taipei there was also a lack of insight into the nature of mental illness and what the services could do for the patients, as well as a fear that they might be subjected to experimentation. In Washington the psychiatrists had to travel one or two hours to go from the FRC to the facility where the patients were interviewed.

Three centres — Cali, Ibadan, and Prague — complained of an irregular flow of patients contacting the facilities. In Prague and Aarhus the large number of patients made sampling necessary.

Other factors that tended to make work more difficult were intra-country migration in Ibadan due to war, and intra-town migration in Moscow, caused by intensive housing construction. Stability of staff no doubt helped the progress of the task of interviewing, and only the Prague FRC mentioned frequent changes of staff as a difficulty. A great advantage was that every Centre retained the same chief collaborating investigator throughout the study.

3.3 Data Processing

3.3.1 General considerations of data processing

When planning of the project first began at a meeting of collaborating investigators held in Geneva in 1966, a subcommittee discussed general problems of data processing and recommended that Headquarters should work out an operational set-up specifying the types of personnel and time required for the reception, scrutiny, coding, and analysis of the enormous amount of data expected to be generated by the study.

Further recommendations of the subcommittee were that the data collected must be uniform; in the forms and schedules, items under any subheading should be mutually exclusive and definitions and instructions for categorization should be given. Only translatable terms should be used and precise local equivalents should be formulated. Only items that would have a similar relevance in the majority of centres were to be included. In order to ensure uniform coding both of simple ratings and of narrative accounts, this

task had to be carried out at Headquarters.

The first tabulations were to be kept simple and were to be aimed at determining the feasibility of continuing the study. It was felt that, since the standards of data collection would vary among centres, in the first year a norm of what could be done by all FRCs would be established to which they could then be asked to adhere. All subsequent data analyses were to be strictly related to the aims of the study and were to proceed from the simple to the more complex.

3.3.2 Data processing

Funds were made available for employing statisticians in the Centres in developing countries. Their first duty was to work out the operational details for local data collection and to train the FRC clerks in their duties. In order to ensure that the study data produced in their Centres were complete and as accurate as possible, the statisticians had to scrutinize all forms and schedules to see that every item was entered, especially where parts of the data were collected by different members of the FRC staff. The importance of their work soon became apparent, since considerable difficulty was experienced with data processing whenever the schedules had not been checked and missing items had not been inserted at the time. In the IPSS whenever information related to the patient's sex, age, or diagnosis, or to whether the type of interview was single, simultaneous, or consecutive was missing, letters of enquiry were sent although this meant that amendments had to be made at all stages of the work. It would have taken too much time to adopt this procedure for individual items, which therefore had to be coded as missing.

Considerable work had to be done in preparation for data handling, processing, and analysis at Headquarters. Staff was trained and arrangements made for the administrative handling and processing of data. Complete code books were developed for Phase 1 and then amended to fit the alterations of instruments made after the preliminary phase of the project. Where this was not possible, correspondence codes had to be established. Similarly, provision had to be made for the linkage with follow-up data. Data processing and analyses at Headquarters was done in close collaboration with WHO technical units of Health Statistical Methodology and Data Processing.

When the main part of the study began in April 1968, all the preliminary preparations had to be brought into full action. It had been arranged that the schedules should be sent in batches from the FRCs, each parcel containing all the documents relative to any given patient. Whenever a batch of schedules was despatched, a separate form was sent to Headquarters giving the patient's name and stating the number of each type of document included. This enabled Headquarters to keep track of schedules despatched from Centres and to take appropriate action if they did not arrive.

After the contents of the parcels of schedules had been checked against the forms received and the Centre had been notified either of their receipt or of any missing documents, a WHO serial number in the project was assigned to each patient and the date of first interview and of arrival of the schedules was recorded. The front page of each schedule was examined to make

sure that the patient's identification data (age, sex, and marital status) were the same on each schedule and that the patient had in fact passed the screening procedure.

The coding, editing and checking of the schedules was done at Headquarters. Although in most Centres the collaborating investigators and other members of the research team devoted a great deal of time and effort to producing data that were as free of errors and omissions as possible, full scrutiny was given all information received at Headquarters. Coding and scrutinizing the data was very time consuming. The amount of time which had to be spent on coding can best be seen from an example. For each patient 19 items had first to be coded identifying the patient, the interviewer, the rater, the Phase of the study, and the type of interview and schedule; these were automatically reproduced on every punch card relating to the patient. Another 10 items concerning the patient's demographic characteristics also had to be coded. It is estimated that, when trained, the coder required six working weeks to complete only this part of the schedule.

The coder then had to scrutinize the ratings in the PSE schedule and verify that a code was entered for every item. As stated above, in Phase 2 this part of the coding was punched directly from the schedule. In all, there were approximately 20,000 80-column punch cards containing the material from the main phase of the project. Of these, 6,000 had to be prepared on hand-written punching slips. It was clearly impossible for there to have been no errors in reproducing more than one million signs on the print-outs.

If scrutiny were restricted to the first 30 items (patient identification, etc.) it would still have been necessary to examine 400,000 signs. If the scrutinizer were allowed to make only one-tenth of the errors normally allowed to the puncher, he would have been allowed to pass 400 mistakes in this section of the data alone. In any large scale data-processing operation the skill of the statistician has to be brought into play to compensate in so far as possible at the tabulation stage for errors not previously detected.

3.3.3 Distribution of data to the Centres

In order for the data obtained from the study to be utilized to the fullest extent, the coded schedules from Phase 1 were all returned to the FRCs. For the Phase 2 data, sets of punch cards for their own data were sent to five of the nine Centres. Four Centres were supplied with cards or tape for more than half the patients in the study so that they could experiment with data analysis, and two Centres had tapes containing all the data from the PSE, Diagnostic Assessment Form and other related instruments for the entire 1,202 patients and also for the inter-centre reliability interviews. These Centres have collaborated with Headquarters in testing methods of data analysis.

In addition, all Centres received copies of the statistical tables prepared for meetings of investigators or exchanges of visits. Many of these analyses were made for individual centres and are not presented in this volume.

3.3.4 General conclusions on data processing

The comments offered here are preliminary thoughts only; final opinions can only be given once the whole process of data analysis has been completed. Nevertheless, they may be of some value to those embarking on international projects of similar size and complexity.

It is desirable that the logistics of data collection and processing be given the same importance at the planning stage of a project as is accorded to the development, evaluation, and standardization of the research instruments, the establishment of centres, and the training of personnel. Data of even better quality might have been obtained had the same effort been put into instructing the local statistical and clerical staff on their expected functions in the project as was devoted to training the psychiatrists.

It is essential to pre-test the elementary parts of data collection, such as reception, recording, and coding, and to determine from the resulting time assessment the number of staff required at various levels. Otherwise, either the work will fall behind or the time of professional staff will be taken up in doing tasks that could be done by research assistants with less training. If, however, funds are not available to provide the necessary staff, the amount of data should be reduced to levels that can be handled by ruthlessly eliminating any item not vital to the aims of the project.

Similar considerations apply to the planning of data analysis; the minimal tabulations and analyses necessary to demonstrate whether the aim of the project has been accomplished should be specified in the original study plan and then rigidly adhered to. If this is done, intermediate analyses can be devoted to testing and hence improving the quality of the work being done, rather than anticipating, on the basis of partial data, results that may not be confirmed after all the data are analysed. As an example of the success of this policy, one such tabulation showed the Centres the percentage of items in the PSE schedule for which no rating had been entered. The result was a considerable improvement for the rest of the study period.

A hierarchy should therefore be established in data analysis at the outset to distinguish between directional and final analyses. At this time too, plans for feedback to the Centres should be formulated so that the FRCs will know what they can expect to receive from Headquarters.

Of course, in almost any research project problems will be raised and data generated that are not strictly germane to the original aims, and other interesting ways of analysing data will suggest themselves. However, it is preferable to put these aside and return to them later.

Finally, a word should be said about the difficulties of fully involving the collaborating investigators in the planning and assessment of data processing. Ideally, the preliminary results of data analysis in the IPSS should have been sent to the collaborating investigators well in advance of meetings to give them an opportunity to study the techniques and formulate suggestions for improvements. Unfortunately, the large amount of data, limited staff, delays in computer processing, postal delays, and other technical difficulties made this impossible. As there was far too little time

available during the meetings of investigators and exchanges of visits for full study of the tabulations, it was not possible for the collaborating investigators to participate as much as Headquarters would have desired in the planning and assessment of data processing. It is hoped that in similar projects ways can be found to provide opportunities for fuller involvement of the collaborating investigators in this aspect of the research.

DESCRIPTION OF THE FIELD RESEARCH CENTRES

4.1 Introduction

The differences between the cultures in which this study was carried out are great and it would exceed the framework of this Report to describe them in detail. First we shall give a brief narrative description of each of the Field Research Centres (FRCs) and its catchment area in the context of the country in which it is located. A tabular presentation of some important statistical information concerning all the Centres follows.

4.2 Description of the FRCs and their catchment areas4.2.1 Aarhus

Denmark, an independent kingdom since ancient times, has never in historical times been subject to much migration. The population, now comprising five million inhabitants, is thus characterized by stability and homogeneity. More than half the population lives in rural areas. The country, consisting of one peninsula (Jutland) and a great number of islands, has no other natural resources than the soil, which is fit for agriculture, and the surrounding sea. Until the 20th century the majority of the inhabitants was therefore occupied in agriculture, fishing, shipping, and trade. As the population has increased, the economy has become vitally dependent on import and export. The export of industrial products has acquired first rank within the national economy during the past few decades.

A liberal and progressive constitution was adopted in 1849. From 1864 to 1940 Denmark was preserved from participation in wars. Both these facts, no doubt, contributed to adequate development of the social and health systems. Insurance systems in these fields developed gradually on a private basis and finally in 1933 became compulsory for the majority of the population. A considerable increase in pensions for old age and disabled pensioners, introduced in 1955, contributed greatly to the social and economic homogeneity of the population. Today the population is economically very uniform: there are few who are very poor and even fewer who are very wealthy. Approximately 40 per cent of the population is employed. Of these one-third are salaried employees, one-half are labourers, and one-sixth are self-employed. One-ninth of the population has a university degree.

The great majority of hospitals are public, general hospitals being governed by cities or counties. Psychiatric hospitals, with 25% of all hospital beds, are governed mainly by the state. During the past few decades an increasing proportion of the work of the hospitals has been performed on an outpatient basis, particularly in psychiatry.

The birth rate is relatively low, and so is mortality. Most infectious diseases have been eradicated. Geriatric diseases are becoming an important

concern in the field of health. The suicide rate is traditionally said to be high, which probably reflects a good registration system rather than a true difference in incidence.

The IPSS FRC is located in the city of Aarhus, which is the second largest city in Denmark, with about 200,000 inhabitants. The Centre is situated in the Aarhus Psychiatric Hospital, a State hospital that serves a population of 832,000 and is the only psychiatric institution within the area, except for two small psychiatric wards in general hospitals (and institutions for the mentally retarded). The hospital was founded in 1852; in 1943 it became affiliated with the University of Aarhus and since then staff and facilities for research have developed rapidly. In 1943 there were seven doctors; in 1971, there were sixty. The staff totals about 900. The psychiatric hospital has approximately 900 inpatients, about 120 day patients and more than 400 long-term patients in nursing homes that are branches of the hospital. There are outpatient clinics in four towns within the district. There are 2,500 admissions per year and about 20,000 outpatient consultations. All treatment is free.

The psychiatric hospital works in close cooperation with the three county general hospitals that serve as university hospitals. It is the clinical facility of the psychiatric department of the University of Aarhus Medical School, and undergraduate and postgraduate teaching is carried out there. Special departments of forensic psychiatry, brain pathology, cytogenetics, and other research units are part of the hospital.

The offices of the FRC are located within the Institute of Psychiatric Demography, which is a part of the hospital and contains a nationwide Psychiatric Register to which all admissions to and discharges from psychiatric institutions are reported.

4.2.2 Agra

Agra, the city of Taj, at one time a bastion of the Moghul empire, is situated in the fertile land between the two sacred rivers Ganga and Jamuna. The region is densely populated and has always been one of the epicentres of Indian history. One of the most sacred places in India is the nearby town of Mathura, which is the birthplace of Lord Krishna who presented the philosophy of the Gita during the battle fought around Delhi described in the epic Mahabharata. Initially a Hindu stronghold, Agra was overrun by the Afghans and became one of the capitals of that important sultanate. When the Moghuls defeated the Afghans they made Agra the fabulous capital of their huge empire. The site of many historic battles, Agra today is an important cultural and economic centre in the central Indian plain.

The psychiatric facilities in the country are scarce and the catchment area of the FRC is correspondingly huge. It stretches over more than 50,000 square kilometers and contains more than 17,000,000 people. The population in the area is increasing rapidly both because of a high natural increase (16.66%) and because of the decreasing death rate. Forty percent of the population is in the youngest age group (0-14). Approximately 10% of this age group is married or widowed. Marriages are usually very stable and monogamous in the Hindu population. Polygamy is still permissible among the

Muslims but is becoming more and more infrequent. Extended families, which were in predominance only a few decades ago, are giving way to nuclear-type families, particularly in the urban areas.

Most of the population is engaged in agriculture, and 80% of the population lives in villages. People are intensely religious and the caste system with more than 70 observed castes makes for a very complicated sociological structure, particularly now when traditional rules are becoming less strict and new value systems are finding their way into the classical setting. Several very different dialects and at least two scripts are used.

Some of the communicable diseases that have been the leading causes of death have recently declined in importance. Malaria and smallpox are examples of these. Tuberculosis, deficiency diseases, and diseases connected with insufficient environmental sanitation still attract major portions of the health budget. Extremes of climate, enormous distances, economic difficulties, and other ecological factors aggravate the health care situation and in particular care for the mentally ill, which is still a low priority field of medical care.

The FRC is located in the Agra Mental Hospital, which has 718 beds. A Medical Superintendent and ten Medical Officers look after the 800 annual inpatient admissions and also run an extensive outpatient service with more than 5,000 visits a year. Undergraduate and postgraduate teaching is carried out in the hospital, and students come not only from the Agra Medical College but also from five other colleges that have no departments of psychiatry. Interviewing rooms with one-way screen, audio- and videotaping facilities, record offices and other facilities of the FRC are in a new wing of the hospital and serve both for project activities and for postgraduate teaching.

4.2.3 Cali

The city of Cali lies on one side of a 200 kilometer long valley hemmed in by two enormous mountain chains. The primitive and warlike Indians living in the valley were either killed by the Spaniards who conquered the valley in the sixteenth century or absorbed by intermarriage. The few surviving Indians who were not subdued sought refuge in the high mountains and the valley was divided among the conquistadores who thus became a land-based aristocracy. Black slaves were brought in to work the haciendas. Colombia became independent of Spain in the second decade of the nineteenth century and abolished slavery in 1850. Some industry developed but agriculture remained the main economic activity, and Cali is the agricultural capital of Colombia. After the second world war an industrialization programme was put into operation. This programme resulted in internal migration that depleted the farms and caused a population explosion in the towns. Cali grew from a town of about 100,000 inhabitants in 1947 to a teeming city with one million inhabitants in 1970. The influx of population to the towns was particularly strong during the period of unrest and guerrilla warfare, which prevailed in the surrounding mountainous areas for nearly fifteen years.

The population of the catchment area is still very young; 40% of the population is below the age of 15 years. The labour force comprises 34.6% of the total population, and of these one-sixth is unemployed. More than

half the population lives in inadequate housing facilities. It has been estimated that 30,000 dwellings in Cali would need to be destroyed, replaced, or considerably improved to make them fit for occupation.

Almost half of all deaths occur before the age of five. The high infant mortality is accompanied by a high maternal death rate. In the age group of 15-44, as much as 55% of all deaths are violent deaths (including accidents, homicide, and suicide). The health services are not well developed, despite the fact that 71% of people who feel ill seek medical services.

The psychiatric services for the catchment area are provided by two in-patient institutions, one with 20 and the other with 245 beds. The latter is the San Isidro Hospital, which houses the FRC. Outpatient services are provided by the San Isidro Hospital and by a Social Security outpatient department, which between them take care of more than 15,000 outpatient visits a year.

The FRC is located in the San Isidro Hospital, which is a part of the University Medical Centre of the University of Valle. Thirteen psychiatrists work in the hospital, and the total staff is approximately the same as the number of beds. More than 2,000 patients are admitted to the hospital each year and of these about one-half are new admissions.

4.2.4 Ibadan

A small village in 1829, Ibadan today is the second largest city in Africa with more than a million inhabitants. It became a town during the decline of the kingdoms in Western Nigeria when soldiers from the collapsing empires decided to settle in this area, which offered protection because of its geographic characteristics. The economic and political importance of Ibadan has grown since the middle of the nineteenth century, and today it is the capital of Western Nigeria and one of the economically most important towns on the continent.

Abeokuta, the second city in the catchment area, came into being in 1830 when people from five different kingdoms gathered under Olumno Rock seeking shelter from the raging tribal wars.

The catchment area covers the very heart of the Yoruba country. The main language is Yoruba but a number of other dialects are spoken. The traditional art is world renowned. The real basis of the Yoruba economy is farming, but the farmers are city dwellers who live as far as 15 miles from the farm belt. The major religions are Christianity and Islam, although in both Ibadan and Abeokuta the majority of the population is Muslim. Both Islam and Christianity are somewhat dissimilar from the traditional concept of these religions and sometimes incorporate elements of animistic religions.

It is difficult to assess the morbidity and mortality rates for children in this catchment area. Birth registration is not universally practised. Literacy is not high. Reporting of death and disease, particularly in children, is probably also somewhat unreliable. From the information available, it appears that infectious diseases cause most deaths. Accidents come surprisingly high among the causes of death (8.1% of all deaths).

The numbers of health staff available are low. More than 17,000 inhabitants rely on one medical practitioner. Health workers other than doctors

are widely used and carry a substantial part of the health care burden. There are many traditional healers, who treat a large number of patients, particularly those with psychiatric disorders.

Institutions providing medical care are expanding constantly. Still, at present there are only 1.2 beds per thousand population, which is more than 10 times lower than the corresponding figure in some developed countries, for example, in the USSR.

Psychiatric services in the catchment area consist of the Department of Psychiatry of the University of Ibadan (10 beds), the Abeokuta Mental Hospital (100 beds), and the Lantoro Mental Institution (350 beds). Village units, which were an ingenious answer to the problem of providing economical and effective service in a culture-specific way, have grown from the original one in Aro Village to four. These units play an important role in the treatment and rehabilitation process. Patients receive treatment as they would in a modern hospital; yet they are not taken out of the context of the culture in which they will live after the treatment is completed.

Undergraduate and postgraduate training is organized in such a way that students see and work in each of these types of service.

The administrative coordination of the FRC was carried out from the Department of Psychiatry in Ibadan. Clinical facilities and staff from this Department, from the Abeokuta Mental Hospital, and from the Aro Village collaborated in the conduct of the study. Videotape and audiotape facilities were located in Ibadan. Interviewing rooms with a one-way screen were provided in both the Department of Psychiatry and in the Abeokuta Mental Hospital.

4.2.5 London

"How to describe a city? Even for an old inhabitant it is impossible; one can present only a simplified plan, taking a house here, a park there as symbols of the whole. If I were trying to describe London to a foreigner, I might take Trafalgar Square and Piccadilly Circus, the Strand and Fleet Street, the grim wastes of Queen Victoria Street and Tottenham Court Road, villages like Chelsea and Clapham and Highgate struggling for individual existence"

Camberwell, in south-east London, where one-third of the patients in the London series lived, is not Graham Greene territory. It was a market garden for London in 1800, with a population of less than 10,000. Then came the tremendous waves of immigration during the industrial revolution, and by 1910 there were 210,000 inhabitants. Since then, Camberwell has been gently declining and the population is now only 170,000. The area retains a markedly Victorian flavour: respectable, old-fashioned, and a little drab. It is solidly working-class, with rather less geographical mobility than the rest of conurban London, a higher proportion of West Indians (4.8% compared with 0.6% for England and Wales) and relatively more unskilled manual workers.

The local social and medical services are very comprehensive and available to all. Mortality follows the usual patterns of urban areas in industrially developed countries. The psychiatric services for the catchment

area of Camberwell are provided by the administrative complex formed by the Maudsley Hospital, Bethlem Royal Hospital, and St. Francis Hospital. Together, these form a postgraduate teaching institution with a wide range of special units for inpatients and outpatients. Patients are also drawn from a much wider geographical area, including metropolitan London and to a lesser extent the rest of the United Kingdom. The FRC is located in the Medical Research Council's Social Psychiatry Unit at the Institute of Psychiatry, which is on the same site as the Maudsley Hospital. A psychiatric case register covering all Camberwell residents who make contact with psychiatric services has been operative since 1964 and forms the basis for a substantial part of the Unit's work on the planning and evaluation of community psychiatric services.

4.2.6 Moscow

As an urban settlement, Moscow came into being not less than a thousand years ago. The favourable combination of its geographical position with rich natural resources, the availability of water and land trade routes, promoted the development of handicrafts and trade as well as the growth of the city itself. By the sixteenth century Moscow had developed into one of the largest cities in the world. In the seventeenth century 200,000 people lived within the boundaries of the "Earthen City" and by the beginning of the twentieth century the territory of the city had grown to 228 square kilometers and its population exceeded two million.

At present the area of Moscow is equal to 900 square kilometers and the population, according to the 1970 census, is 7,061,000. About 60% of the population is employed, and approximately 17% is made up of old age pensioners, retired people, and invalids.

The birth rate in Moscow in 1966 was 11 per 1,000 population, and the death rate was 8.8 per 1,000. Thus, the rate of population growth equalled 2.2 per 1,000. More than 30% of families are households consisting of three members, and slightly less than 30% of families have four to five members.

Moscow's network of medical facilities has been expanding constantly. While in 1914 Moscow had 50 hospitals with 10,600 beds, by 1963 there were 300 hospitals and 77,000 beds. All medical facilities are maintained by the state, treatment is free of charge, and drugs are free for certain groups of patients, including schizophrenic and epileptic patients.

Psychiatric care is rendered to the population of Moscow through a number of specialized facilities. Eighteen district and inter-district Psychoneurological Outpatient Dispensaries perform outpatient observation and treatment of mental patients, as well as provide different kinds of social assistance. All the Dispensaries and Mental Hospitals have sheltered workshops where much attention is paid to work therapy and social and work rehabilitation. In the system of the Ministry of Social Welfare there are several shops for mental invalids.

Special psychoneurological polyclinics take care of children under 15 years of age; adolescents 16-18 years old are served at the Outpatient Dispensaries. This group of patients receives consultations at and is ad-

mitted for inpatient treatment to the City Psychoneurological Dispensary (with the Inpatient division) for children and adolescents.

In the FRC the majority of staff belongs to the Epidemiological Department of the Institute of Psychiatry of the Academy of Medical Sciences of the USSR. The Centre is headed by the Vice Director of the Institute and the Chief of the Epidemiological Department. In its work the Centre enjoys the full support of the Institute and the assistance of the OPD in the Kievskiy raion, that is, one of the Institute's bases.

4.2.7 Taipei

Taipei is the largest city of Taiwan. Situated in the northern part of the island, it is Taiwan's political, economic, and cultural centre. The city is located in a basin ringed by low mountains, and three rivers, the Tanshui, Keelung, and Hsintien, provide outlets to the Taiwan Straits to the West and the Pacific Ocean to the East.

The metropolis and the neighbouring counties of Taipei, including a few small cities and towns, are included in the catchment area of the Taipei FRC. Many government offices, major universities, culture centres, business and industrial offices, entertainment and tourist centres are located inside the metropolis, and these are mingled with residential areas and a few districts of low income housing and shacks of native workers and migrant labourers. Due to the pressure of rapid population increase and industrial development, the adjacent rural land is rapidly disappearing and the urban:rural ratio (64.8 : 35.2) is changing accordingly. The population has increased about four times or more in the last fifteen years due to a high rate of natural increase and the immigration of labourers, job seekers of various categories, refugees, and retired soldiers.

The living conditions are generally good in spite of housing shortages and strained public services. Because of the large percentage of youth in the population -- approximately 40% under the age of fifteen -- the public schools for compulsory education (lasting nine years) are extremely overcrowded. The traditional large family household is still prevalent, but owing to the increased mobility of the population coupled with the housing shortage, the average household is becoming smaller (5.6 persons) and the average number of children per family is 3.2.

Precise statistics on occupational distribution of the population and unemployment figures are not available, but unemployment or under-employment is quite a serious problem. There are also no precise statistics for income of the inhabitants but the average income is low. For example, the salary of a public school teacher is equivalent to \$50 a month. Thus, many members of the family have to work to contribute to the family living.

General health conditions have improved greatly and are still improving; the infant mortality rate is estimated at about 20-30 per thousand live births and the maternal mortality rate is .43 per thousand live and still births. Acute infectious diseases or malnutrition have given way to other diseases as leading causes of death in the last decade; cerebrovascular disease and cancer are the first and second leading causes of death, and are followed by pneumonia, accidents, and tuberculosis, in that order.

Health personnel are reasonably adequate. For example, there are about 2,200 doctors available, i.e., one doctor per 1,100 inhabitants. There are five general hospitals and a large number of small private hospitals in the city with 1.9 beds for 1,000 inhabitants.

The Taipei FRC is situated in the Department of Neurology and Psychiatry of the University Hospital. The hospital together with its grounds occupies about twenty acres in the city of Taipei. It is equipped with 900 beds and served by a work force of about 1,500, which includes teaching staff and both professional and non-professional employees. The Department of Psychiatry has about 50 beds; there are 20 day-hospital patients and about 100 outpatients each day. Undergraduate and postgraduate teaching are carried out in the Department.

In addition to the Department of Psychiatry of the University Hospital, one charity mental hospital and two small private mental hospitals with a total bed capacity of 356 served as psychiatric facilities for screening patients in the present study. These facilities represent about 23% of the total psychiatric beds in the city. The number of acute admissions to these affiliated facilities amounted to about half or more of the total psychiatric admissions in Taipei.

4.2.8 Washington

Prince Georges County, Maryland, the catchment area for the Washington FRC, stretches from the south-east of Washington, along the eastern bank of the Potomac River, north-west and eastward, halfway to Baltimore.

Because of its proximity to Washington, the County has a large suburban population, which makes up 80% of the population of the County but occupies only the 20% of the County's land area immediately adjacent to Washington. The remainder of the County consists primarily of rural areas and many small towns, although with the rapid increase of suburban development the size of the rural area is shrinking.

Partly because of the general desirability of the County as a place in which to live and because of the expansion of federal employment, there has been a rapid population growth through immigration from other parts of the United States. A threefold population increase has taken place since 1950, with an additional twofold increase predicted by 1980. Since employment in federal or local government accounts for about 35% of the work force and since many of these positions are temporary, there is more mobility of the suburban population both into and out of the County than in many other parts of the United States. The high percentage of government employment assures at least a moderate level of income so that it has been possible through local taxation to support the facilities necessary to accommodate the population growth.

The inhabitants of Prince Georges County have a mean age of about 25 years, younger than that in many areas in the United States. This is probably attributable to the tendency of younger, more mobile people to move into the area to take advantage of the jobs available and to raise their children there.

Health conditions in the County are good. The leading causes of death

are heart disease, cancer, cerebrovascular disease, accidents, and influenza and pneumonia, in that order. Suicide ranks 10th as a cause of mortality.

Medical facilities serving the population of the County are located in the County itself and in surrounding parts of Maryland and Washington D.C. In contrast to several other centres participating in the IPSS, the use of health care facilities in this catchment area is not restricted by specific geographic boundaries. The majority of the residents use one of the five large general hospitals serving the part of the larger metropolitan area in and around Prince Georges County. Most medical care is provided by private practitioners and usually paid through comprehensive health insurance. Low-cost and free clinics are also available.

The Washington FRC consists of two separate parts: (1) the research team, a section of the research division of the National Institute of Mental Health, and (2) the three psychiatric facilities that provide the majority of psychiatric care to residents of the Prince Georges County Catchment Area.

Inpatient psychiatric care in Prince Georges County is provided by Spring Grove State Hospital, a large state hospital actually located outside of the boundaries of the County; Prince Georges General Hospital; and Cafritz General Hospital. Since all three facilities collaborated in the IPSS, a large percentage of all County patients requiring hospitalization for psychiatric care was screened for the study. Outpatient facilities in the catchment area consist of private psychiatrists, county alcoholic clinics, a county mental health clinic, a state hospital followup clinic, and the community mental health study centre of the National Institute of Mental Health. The sparsity of facilities available for care of those patients who do not require full time hospitalization but still require outpatient treatment or part-time hospitalization, is a problem in Prince Georges County. To some extent, the development of community mental health centres is overcoming this deficiency, but at the time of the IPSS the availability of outpatient and intermediate levels of care was still a problem.

4.2.9 Prague

Prague has been the capital of Bohemia since the ninth century. The first traces of people living in the area date back 250,000 years. As early as the third millenium B.C. the Prague area was permanently inhabited by people cultivating the soil, and it has never been a waste land since that time. The Slavs came to Prague in the 5th and 6th centuries A.D. and founded the Prague Castle, which became the cultural and spiritual centre of the country.

In the 14th century, under King Charles IV, Prague became the capital of the "Holy Empire of the German Nation". In 1348 the first University of central Europe was founded in Prague. At this time, Prague was the third largest city in Europe, after Constantinople and Rome.

After the overthrow of the Austro-Hungarian Empire in 1918, Prague became the capital of Czechoslovakia. During the last hundred years the population of the city has increased fourfold and by the end of 1967 it had 1,102,000 inhabitants. Population growth in Prague is under strict control, as it has been estimated that any increase in the population beyond 1.5

million would cause severe technical and hygienic complications. Since 1959, the population growth of Prague has been increased only by migration.

In Prague, there is a small proportion of children and a large proportion of old people. In 1967, 17% of men and 22.7% of women were age 60 or over. Such a large percentage of old people creates many difficulties in the field of social and health care.

There has been a great manpower shortage in the city recently. Almost all men and three-quarters of women in the 15-59 age bracket are economically active. Migration as a possible source of manpower is limited by the great lack of housing.

In recent years there has been a decrease in the number of deaths caused by infectious diseases and an increase in the mortality rates for cardiovascular disease, neoplasm, cerebrovascular disease, and suicide. The mortality rate for suicide (34.3/100,000) is more than twice that for tuberculosis (12.3/100,000).

Prague itself is divided into ten municipal districts. The organization of the health service corresponds to the administrative structure. The majority of the city's general practitioners (91% in 1967) work at polyclinics together with specialists.

Psychiatric beds (including those in the mental hospital) represent 19.2% of hospital beds. The number of psychiatric beds amounts to 16.8 per 10,000 inhabitants.

Since 1950, the entire country has been divided into new catchment areas for several mental hospitals. From 1950 to 1963 the catchment area of the Mental Hospital at Bohnice included Prague and some of its surroundings. Since 1963 it has served only residents of Prague. The Mental Hospital at Bohnice was planned as the main psychiatric facility for Prague and it has served this purpose for over 60 years. It consists of more than 30 buildings. Most of the pavillions were built before the First World War.

The Psychiatric Research Institute at Bohnice started functioning on 1 January 1961 in two pavillions taken over from the Mental Hospital. The institutions are independent units with different headquarters. The Institute has had no special catchment area in recent years. The patients included in this study were mostly transferred from the Mental Hospital.

The Prague FRC is located in the Psychiatric Research Institute. Administratively and financially it is independent of the Mental Hospital, being financed directly by the Research Institutes Division of the Czech Socialist Republic Ministry of Health, while the latter is financed by the Health Department Division of the Prague Municipal National Committee. The Director of the Institute is in charge of all its scientific programmes. A Scientific Council (composed of heads of psychiatric departments of medical schools, the Postgraduate Institute of Medicine, the Neurological Department of the medical school and the Physiological Institute of the Academy of Sciences of the Czech Socialist Republic, and the Superintendent of the Mental Hospital) performs an advisory and consulting function to the Institute.

4.3 A Summary Comparison of some Demographic and Medical Statistics of the FRCs

Table 4.1 summarizes the vital statistics and demographic characteristics for the nine FRCs. Examination of the table indicates that the objective of selecting Centres that serve catchment areas with a variety of sociocultural backgrounds was indeed met. The size of the catchment areas ranged from 52,076 square kilometers in Agra to 14 square kilometers in London. The population served by the catchment areas varied from a high of 17,340,000 in Agra to 172,000 in London. Population was most densely concentrated in London (12,180/square kilometer) and least dense in the Aarhus catchment area (62/square kilometer). In only two Centres, Agra and Ibadan, did a minority of the population of the catchment area live in an urban area.

The rate of natural increase of population varied between 3.1% in Cali and minus 2.9% in Prague (despite the fact that the Prague catchment area had the highest immigration rate of all the Centres). Infant mortality rate was by far the highest in Agra (186/1,000 live births); Cali was second highest with one-third the rate of Agra. The remaining Centres all had a rate of approximately 20/1,000.

The age distribution of the population was similar in Aarhus, London, Moscow, and Prague, on the one hand, and in the four Centres in the developing countries, on the other hand. In all of the latter FRCs there was a relatively high percentage of young people (0-14), and the same was true of the Washington catchment area. The average number of persons per household ranged from 8.4 in Cali to 2.6 in Prague.

From the list of the 5 most frequent causes of death it can be seen that, in all centres but two, people die primarily of cardiovascular or cerebrovascular disease and cancer. In Agra and Ibadan, however, infectious diseases are the leading causes of death.

The differences outlined in this table must be borne in mind when evaluating the data produced by this study. For example, when length of hospitalization for schizophrenic patients is compared across centres, it must be remembered that in FRCs with a great lack of psychiatric beds, there may be great pressure to discharge patients as soon as this is at all possible. Similar factors must also be considered when evaluating results of followup studies. For example, in order to use employment as one criterion for assessing readjustment to the community, it is necessary to know what the overall employment rates in the catchment area are.

However, it must be emphasized that the data presented here are not in any sense intended to serve as a complete basis for a study of sociocultural variables. Such a study would require a complete sociocultural investigation of the catchment areas. It is hoped that such investigations will take place in the future.

TABLE 4.1 A COMPARISON OF SOME DEMOGRAPHIC AND MEDICAL CHARACTERISTICS OF THE FRCS

	Aarhus	Agra	Cali	Ibadan	London	Moscow	Taipei	Washington	Prague
Catchment Area									
Area (km ²)	13,000	52,076	352	46,464	14	886	616	1,246	190
Population (thousand)	832	17,340	808	4,915	172	7,061	2,322	638	1,034
Population density (per km ²)	62	333	2,292	106	12,180	7,970	3,764	512	5,421
Population, urban area (%)	76.3	20.5	96.9	37.8	100.0	100.0	64.8	89.0	100.0
Vital Statistics									
Natural increase rate (%)	0.7	1.7	3.1	2.0	0.2	0.2	2.4	1.8	-2.9
Birth rate (per 1000 population)	17.1	41.5	38.5	40.0	16.8	11.0	28.8	22.9	10.4
Death rate (per 1000 population)	10.1	24.9	8.0	20.0	11.7	8.8	4.6	5.2	13.3
Infant mortality rate (per 1000 live births)	15	186	56	(24) ^a	18	22	(12) ^b	16.8	24
Net immigration rate (%)	0.0	(c)	3.2	*	-0.5	(c)	1.6	*	3.8
Demographic Characteristics									
Age distribution (%) male									
0-14	25	39	48	34	23	24	39	33	17
15-44	42	43	49	54	43	51	43	49	44
45-64	23	14	11	9	25	20	16	15	29
65+	10	4	2	3	9	5	2	3	10
Total	100	100	100	100	100	100	100	100	100
Age distribution (%) female									
0-14	23	41	43	35	21	17	41	31	15
15-44	41	43	43	53	40	48	45	48	40
45-64	23	12	11	9	25	26	11	16	30
65+	12	4	3	3	14	9	3	5	15
Total	100	100	100	100	100	100	100	100	100
Median age (years) male	33.9	22.7	17.4	23.9	33.8	30.3	22.7	25.9	37.5
female	34.8	21.3	19.6	23.5	36.8	35.6	21.0	27.3	41.3
Married population (%) male	49	72	29)	*	68	70	48	46	55
female	48	80	29)	*	60	49	57	45	48

* Not available.

a The neo-natal mortality rate refers only to deaths of babies born in hospital.

b The infant mortality rate was reported to be 11.7 in Taipei itself and 25.7 in the neighbouring port of Keelung. However, these rates are not comparable with those in the other Centres, because fetal deaths and those of infants dying up to three days before registration are not included, nor are illegitimate births. In the view of United Nations experts the rate may well be about three times the one quoted. Nevertheless there has been considerable reduction in the rate of recent years.

c Insignificant.

TABLE 4.1 (continued) A COMPARISON OF SOME DEMOGRAPHIC AND MEDICAL CHARACTERISTICS OF THE FRCS

	Aarhus	Agra	Cali	Ibadan	London	Moscow	Taipei	Washington	Prague	
									M	F
Occupational Characteristics										
Unskilled workers (%)	33	8	*	20	12	*	28	4	20	34
Professionals (%)	17	2	*	4	3	*	9	20	23	5
Household Characteristics										
Number of persons per household	2.8	5.7 ^d	8.4	*	3.4	*	5.6	3.4		2.6
One person household (%)	24.3	*	12.5	*	20.3	*	13.1	15.3		22.3
Mortality										
Five leading causes of death										
1st	Heart Diseases	Fevers	M: Heart F: Cancer	Tetanus	Ischaemic Heart Disease	CVD	CVD	Heart Diseases	CVD	
2nd	Cancer	TB	M: Cancer F: Heart	Pneumonia	CVD	Cancer	Cancer	Cancer	Cancer	
3rd	Apoplexia, Senility	Diarrhoea, Dysentery	M: TB F: TB	Malaria	Pneumonia	*	Pneumonia	Cerebrovascular disease	Injuries, Poisoning and Violence	
4th	Accidents	Deficiency Diseases	M: Vascular lesions of CNS F: Vascular lesions of CNS	Measles	Cancer (digestive organ & Peritoneum)	*	Accidents other than motor vehicle	Accidents	Respiratory Diseases	
5th	Respiratory Diseases	Cancer, etc.	M: Other accidents F: Delivery and complications of pregnancy	Dysentery	Bronchitis, Emphysema & Asthma	*	Respiratory TB	Influenza and Penumonia	Infective and Parasitic	

* Not available.

^d Agra only. 4.5 for Kanpur area included in the study.

TABLE 4.1 (continued) A COMPARISON OF SOME DEMOGRAPHIC AND MEDICAL CHARACTERISTICS OF THE FRCS

	Aarhus	Agra	Cali	Ibadan	London	Moscow	Taipei	Washington	Prague
Health Services (general)									
Non-psychiatric Hospitals									
Number of doctors	518	*	88	285	124	e	528	351	562
Number of other hospital staff	7,677	1,420 ^f	898	5,177	871	e	7,462 ^g	1,615	8,184 ^g
Number of beds	4,825	5,000	824	5,527	1,009	77,000	4,765	2,807 ^h	10,502
Beds (per 1000 population)	6.0	0.3	1.1	1.1	6.5	13.7	1.9	4.4	9.5
Health Services (mental health)									
Psychiatric facilities									
Number of Psychiatric Hospitals	1	1	2	3	3	15	15	1	2
Number of Psychiatric Wards in general hospital	2	2	0	9	2	3	2	6	397
Number of Outpatient clinics	8	2	4	1	6	20	8	3	20
Number of Psychiatrists in private practice	3	3	13	*	Nil	Nil	20	60	Nil
Number of beds in hospital & wards	909	738	265	460	496	15,295	1,560	1,525	2,118
Beds (per 1000 population)	1.0	0.04	0.3	0.1	2.9	2.2	1.9	2.4	2.1

* Not available.

e Moscow. Total number of doctors in hospital and non-hospital practice: 49,309
Total number of other hospital and non-hospital personnel: 97,300

f Agra. Excluding health visitors.

g Taipei, Prague. Nurses only.

h Many suburban residents use Washington, D.C., or Montgomery County Hospitals for general medical non-psychiatric care.

INSTRUMENTS

A brief systematic description of each of the main types of schedule used in the study is given in Table 5.1. For each patient about 1,600 items of information were recorded. In addition to the items from schedules listed in the table, a number of items were coded from narrative accounts. The development of certain of these instruments is described in more detail below.

5.1 Present State Examination Schedule

The Present State Examination (PSE) schedule is a guide to structuring a clinical interview, with the object of obtaining a valid and reliable description of the present mental state of adult patients suffering from one of the functional psychoses or neuroses. The schedule was developed over ten years ago and the first eight editions were tested extensively (Wing et al., 1967; Kendell et al., 1968; Cooper, 1970).

Basically, the schedule is a list of items that systematically covers all the phenomena likely to be considered during the examination of a patient's present mental condition and indicates how they are to be coded. Each symptom is defined in terms of a greater or lesser number of items.

Throughout its development the basic principles of administration of the interview have remained unchanged. For most symptoms a form of questioning is suggested. Theoretically, therefore, it would be possible to carry out the entire interview without deviating at all from the schedule. In practice this very rarely happens, since no two interviews are alike and each examiner must be able to adapt his technique to the situation. The principles for conducting the interview flexibly while preserving a substantial degree of standardization are given elsewhere (Wing et al., 1967), but three of the most important should be emphasized. In the first place, the approach is not that of a questionnaire. The interview remains a clinical one; the examiner is free to cross-question as much as he thinks necessary in order to determine whether a symptom is present or not. Secondly, the period of past time covered is restricted to one month before examination. At earlier stages in the development of the schedule periods of three months and of one week were tried, but the one-month period seemed preferable. If the period were much longer the patient would not easily remember what symptoms had been present; if it were much shorter there was a danger that only short-term fluctuations in symptomatology would be recorded at the expense of a wider view of the clinical picture. Symptoms occurring earlier than one month prior to examination were therefore excluded. Only symptoms mentioned by the patient or observed during the interview were recorded in the PSE. Symptoms mentioned by other informants were recorded in the Psychiatric History, Social Description, Condition

TABLE 5.1 DESCRIPTION

Name and Abbreviation for Schedule	Areas Assessed	Method for completing	Time covered	Source of information	Time necessary for administering
Present State Examination (PSE)	Symptomatology of functional psychiatric disorder: interests; concentration; somatic sympts; irritability; mood; obsessional symptoms; derealization and de-personalization; hallucinations; delusions; behaviour in interview; observed affect; speech; rapport	Clinical interview	Present and one month previously	Patient	45-90 minutes
Psychiatric History (PH)	History of: present and past episode of illness treatment; premorbid personality traits; significant events in life story; psychological adjustment to work and psycho-sexual adjustment; use of alcohol and drugs; legal history	Clinical interview; use of case records; interviews or correspondence with professional informants or patient's friends and relatives	Life until examination	Patient; informants; professionals; case records; others (e.g. records from school, prison, etc.)	45-70 minutes
Social Description (SD)	Social functioning before and during illness; description of residence and household; education; work; marriage and children; patient's childhood setting; sibship and parents; religious activities; social and leisure activities	As in PH	Life until examination	As in PH	30-70 minutes
Physical and Neurological Examination (PNE)	Findings of clinical and laboratory examination; particular emphasis on neurological findings	Examination of patient by medical officer; use of laboratory findings either from history or done specially for this Study	At time of examination	Patient; laboratory in FRC	30 minutes

* These numbers do not include items dealing with information for identifying patient.

INSTRUMENTS USED IN THE IPSS

Administered	Layout	Number and Type of Items	Available in	Date Processing	Type of Output
Psychiatrist	Printed 112 pages. Items printed on right-hand page only, left-hand side left for definitions and examples	360 items * Four point rating scales for 347 items; items are in form of question or of a definition of abnormal behaviour. 13 items of Yes, No, ?, type. Of these 360 items, 235 are for recording reported symptoms; 107 for observed behaviour and 18 for factors affecting quality of information	Chinese Czech Danish English French Hindi Russian Spanish Yoruba	Direct punching from schedules (Protocol key-punch). Examples and comments were written and handled separately	Analysis in terms of 1) ratings on individual items Items aggregated into 2) units of analysis and groups of units of analysis 3) symptoms, syndromes, and diagnostic groups of CATEGO 4) clusters based on similar symptomatology 5) transcripts of interviews
Psychiatrist; other medical officers; psychologists	Printed, 40 pages for recording information; right-hand pages only used, left-hand pages used for instructions and examples	207 items * of four kinds: 1) requiring a narrative account similar to open-ended questions 2) with checklists (pre-coded) 3) dichotomous items 4) rating scales	As in PSE	Transcription to coding sheets for all pre-coded items; content coding for narrative accounts	1) ratings on items 2) 61 units of analysis with dichotomous ratings describing 15 variables
Usually psychiatric social worker or psychologist; can be done by psychiatrist or other medical officer	Mimeographed, 30 pages, both sides of sheets used. Instructions given in a separate 10 page document	152 items * Open-ended questions; checklists and rating scales; semi-open-ended questions	As in PSE	Transcription to coding sheets for pre-coded items; content coding for the narratives	Ratings on items and units of analysis
Medical officer	Printed, 5 page document, both sides of pages used	17 items * relating to physical examination and anthropological measurements; 13 items for laboratory findings; 14 items for neurological findings; descriptions of areas of functioning are given and ratings of dichotomous type provided	English only	Transcription to coding sheets although direct coding from Protocol would be possible for entire neurological examination and large part of physical. Separate analysis of comments and examples or descriptions of findings	Ratings on presence or absence of abnormalities

TABLE 5.1 DESCRIPTION OF
(continued)

Name and Abbreviation for Schedule	Areas Assessed	Method for completing	Time covered	Source of information	Time necessary for administering
Diagnostic Assessment (DA)	Diagnoses; summary of findings, supporting diagnosis	Assessment of material contained in preceding schedules	Present illness	All documents; re-examination of patient if necessary	10 minutes
Screen Form	Demographic information and description of exclusion and inclusion categories for IPSS	Use of available information in FRC and examination of patient	From 5 years before examination until time of examination	Patient; hospital records; professionals; informants	Variable
Condition on Admission Form	Thumbnail sketch of history and pre-admission behaviour	Clinical interview with patient; use of records and interviews of informants	Life until examination	As for PH	10 minutes
Monthly Reports	Description of patients contacting facility	Reviewing FRC record	One month	Hospital records	Variable, depending upon statistical system used in centre
Appendix 5	Description of documents sent	N/A	N/A	Central study file in centre	5 minutes

INSTRUMENTS USED IN THE IPSS

Administered	Layout	Number and Type of Items	Available in	Data Processing	Type of Output
Psychiatrist	Printed, 4 pages, both sides used	(1) 6 open-ended questions requiring summary of findings; (ii) Statement of diagnosis according to system used in centre; (iii) Checklist of ICD categories pertaining to functional psychiatric disorder; (iv) Rating scales for description of prognoses (3 items)	As in PSE	Transcription to coding sheets	Ratings on items (Diagnostic Assessment)
Medical Officer and other FRC staff	Mimeographed, both sides of single sheet	9 identification items and checklists of exclusion and inclusion categories	English Spanish	Processed in centres only; summaries of information content of Screen Forms for patient population provided in Monthly Reports	N/A
Clerk or psychiatric social worker	Mimeographed, 1 page	6 items requiring brief narratives	English Spanish	Done in centres; schedule used only if no other information is available to psychiatrist who is to carry out PSE	N/A
Clerk	Mimeographed, 2 pages	Grids showing number of patients contacting facility; by age, sex and residence and reasons for exclusion of demographically eligible patients	English Russian	By hand	Summary tables for centres corresponding to variables recorded
Clerk	Printed, 1 page, completed in duplicate	Checklist of documents sent to HQ	English	N/A	N/A

on Admission, and Diagnostic Assessment forms. Thirdly, a system of 'cut-off points' was introduced so that if, after exploratory questioning, the items in a particular section did not appear to be eliciting any positive replies, another section could be tried, and so on throughout the interview. On the other hand, if the initial questions concerning a given area were productive or if other clues suggested it, this particular area of psychopathology could be explored in more detail. In general the aim was to conduct a clinical interview with good rapport between interviewer and patient, and it was intended that the various techniques of standardization used should be compatible with this goal.

The seventh edition of the PSE, which incorporated amendments suggested by some of the investigators, was used in Phase 1 of the IPSS. On the basis of their experience during this phase, collaborating investigators made further suggestions about the wording and ordering of items and proposed additions and deletions. Together with the experience of the team working on the US-UK diagnostic project and of members of the Medical Research Council Social Psychiatry Unit in London who had been using the seventh edition in major studies, this enabled an eighth revision to be made. The eighth edition exists in two forms -- a longer version used by the US-UK and MRC research teams, and a somewhat shorter version, omitting some of the items relating to neurotic symptoms and adapted for international use, used in the IPSS. This latter version of the edition is the one referred to throughout this volume as the PSE schedule. Its layout is such that the ratings can be directly transferred to punch cards.

The schedule used in the IPSS thus contains 360 items arranged in 17 sections. There are 94 items in the first 10 sections, covering "non-psychotic" symptoms; 140 in sections 11 and 12, covering disorders of thinking and perception; 1 in section 13, covering insight; 107 in sections 14, 15, and 16, covering observed behaviour; and 18 items in section 17, covering circumstances that might have influenced the quality of information obtained during the interview.

The patient's identifying data and details concerning the interviewer, rater, and type of interview are found on the front sheet. Then follow two pages of instructions to the interviewer. The first section contains questions about treatment and provides a structure for the opening part of the interview, which is introductory and exploratory. The patient is encouraged to describe in his own words the problems that brought him to hospital, thus providing the interviewer with a rateable sample of spontaneous speech and also enabling him to decide where the more formal part of the interview should begin. Ordinarily, he begins formal questioning with section 2, which is concerned with the patient's interests; and this leads naturally to sections 3 (Concentration), 4 (Somatic Symptoms), and 5 (Irritability). Sections 6 and 7 (Slowed Functioning and Depressed Mood) follow. Section 8 is divided into three parts (Muscular Tension and Restlessness, Subjective Anxiety, and Situational Anxiety). Sections 9 (Elated Mood) and 10 (Obsessional Symptoms) complete the "non-psychotic" part of the schedule. Section 11 (Perceptual Anomalies) contains subsections concerning derealization, depersonalization, and various forms of hallucinations.

Section 12, with eleven subsections, deals with delusions and delusional experiences of various kinds. Insight is rated in section 13. Sections 14, 15, and 16 are concerned with abnormalities observed during the interview in the patient's behaviour, affect, or speech. Finally, some general ratings of the quality of the information and the patient's rapport are made in section 17. The names given to the sections were not intended to have diagnostic significance, and all participating psychiatrists were aware that items from any combination of sections could be rated positively. For example, the patient might show signs of both depressed and elated mood (sections 7 and 9) during the same interview.

The items are grouped into sections to facilitate the conduct of the interview, since it would be confusing to the patient and interfere with the natural flow of question and response if they were mixed at random. However, the interviewer is not obliged to follow the order of sections in the schedule. For example, if the patient mentions psychotic symptoms during the introductory part, the interviewer can begin formal questioning at an appropriate point in sections 11 or 12. He can also vary the order in which sections are taken if this is necessary to preserve the patient's interest and attention, or to explore areas of pathology that the patient's remarks suggest might well be productive.

All but 13 of the 360 items can be rated 0, 1, 2, ?, NR (no response), NA (not applicable), or NI (not inquired). The meanings of 0, 1, 2 are indicated in the text of the schedule, since they vary according to the nature of the item. Some items are rated on the basis of frequency of occurrence (e.g., irritability, panic attacks) and some on the basis of severity (e.g. subjective estimate of depressed mood). Most items are rated on a combination of the two (e.g., 1 - occasional or not severe, 2 - continuous or severe).

5.2 Psychiatric History and Social Description Schedules

Whereas the PSE schedule had been developed and tested before the IPSS began, the development of a psychiatric history schedule and a standardized social description was not so advanced.

When the study began, the London and Washington FRCs were already using history schedules developed in the course of their own research projects, and some of the questions in them were used in the first draft of the IPSS Psychiatric and Social History schedule, which included items relating to both psychiatric history and social functioning. This first draft was used in Phase 1. However, since some questions had to be asked by a psychiatrist while others could be rated by a psychologist or social worker, it was decided to produce two separate schedules for use in Phase 2: the Psychiatric History schedule and the Social Description schedule. It was difficult to decide which questions belonged most appropriately to which schedule, but this was determined by putting questions that would normally be asked by the psychiatrist into the Psychiatric History and the rest into the Social Description.

Because these schedules were developed so late it was not possible to

give them the same pre-testing that the PSE had received.

There were also technical difficulties that impeded a thorough testing of reliability. In some centres there was only one person completing both schedules; the proportion of information obtained from the various sources, e.g., from informants, case notes, etc., varied both from centre to centre and from patient to patient; the professional training of the team members completing the schedules was not uniform (some were, for example, psychologists, others psychiatric social workers); the schedules had items requiring narrative accounts, the inter-reliability of which it is difficult to assess.

An assessment of the validity of history or social data is even more difficult than the assessment of validity of the PSE. For example, if a patient claims he was in a certain hospital five years ago his statement can be checked, whereas if he denies ever having been in a psychiatric hospital before the truth of this statement is much harder to assess.

A particular problem in drafting these schedules was to determine what sort of information would be most useful in defining a relationship between social factors and diagnosis. Since there is little definite evidence available on this question, the choice of questions tends to depend on the orientation of the person preparing the questionnaire. For example, a psychiatrist mainly oriented towards the patient as an individual is likely to look for events in his past that may be connected with instability or that are known to cause psychiatric trauma. The sociologist, on the contrary, is more likely to include questions about the patient's general environmental setting, and in the social description he will probably use factors that enable society as a whole rather than the individual patient to be categorized in terms of various characteristics. In addition, there was of course the problem of comparability in assessing socio-demographic variables across cultures. It is worth noting that discussions about the history and social description schedules both before and in the course of the study, often revolved around the decision between a fully precoded questionnaire and a series of open-ended questions. The former is easier to handle but may hide cultural differences; the latter are more likely to discover such differences but present problems for data processing and comparisons.

Another difficulty in developing a suitable questionnaire for use in a transcultural study is that much of the work establishing relationships between, for example, life-events and mental breakdown, has been done in certain types of culture only, and these connexions may not exist elsewhere. It is difficult for the person developing a history schedule to empty his mind of the facts that, in his culture, he has learned to be associated with mental illness. Yet to increase the length of the schedule by the indiscriminate inclusion of items with a possible bearing on mental health would have placed an undue burden on both interviewer and patient. In the questionnaires developed in the IPSS, perhaps too much attention was paid to events and too little to exploring the patient's reaction to them, although this reaction might significantly influence the interpretation of the findings. For example, a period of hospitalization may appear desirable because at least it ensures shelter and some regular meals.

The cross-cultural difficulties involved in interpreting data from these schedules are of course also immense. For example, if the patient's work history includes constant changes of occupation, in one centre these may be regarded as indicative of instability and categorize the patient as irresponsible and a 'drifter', while in another centre such changes may reflect the difficulties of the general economic situation, and the fact that the patient has succeeded in keeping himself employed at all may be very much in his favour. In still other cultural settings a regular change of jobs may be thought desirable.

During Phase 2, the Psychiatric History interview was carried out by a psychiatrist, other medical officer, or psychologist, and the Social Description schedule usually completed by a psychiatric social worker or psychologist. In both schedules the source of information was recorded under the headings of Patient; Informant; Professional Source; Other. The interviewer was allowed to take the items in any order he wished but was urged wherever possible to use the questions in the form proposed. Some questions had pre-coded answers; there were a fair number that had to be answered with a free narrative; and some had rating scales or grids to be filled in by the interviewers.

It was realized at the outset that there would be difficulty in obtaining data about experiences and symptoms in childhood and other events remote in time. While data from such items may not be very valuable for describing patients in this study, they may nonetheless prove useful in analyses directed towards the production of valid instruments, which was one of the major goals of the IPSS.

5.2.1 Description of Psychiatric History schedule used in Phase 2

The following areas were covered in this schedule:

- (1) Onset, symptomatology and course of illness
- (2) Treatment
- (3) History of contacts with the medical services
- (4) Other persons or organizations whom patient has consulted about his psychiatric problems
- (5) Behaviour symptoms at any age
- (6) Patient's premorbid personality traits
- (7) Life history
- (8) Work history
- (9) Psychosexual adjustment

- (10) History of use of alcohol and drugs
- (11) Contacts with the law
- (12) Patient's overall satisfaction with his premorbid life situation
- (13) Assessment of sources of information

The development of the Psychiatric History schedule from the combined document used in Phase 1 was a time-consuming operation, because each version of the document had to be circulated to FRCs and their replies incorporated in the final product. Although long and drawn out, this process was nevertheless rewarding. Thanks to the comments of the collaborating investigators, it was possible to produce a schedule that was applicable in all the Centres and that proved useful in the course of the study.

5.2.2 Description of Social Description schedule used in Phase 2

The following areas were covered by this schedule:

- (1) Residence and household
- (2) Education
- (3) Work activities
- (4) Education and occupation of spouse
- (5) Education and occupation of parents
- (6) Education and occupation of head of current household
- (7) Religion
- (8) Patient's childhood setting
- (9) Marital status
- (10) Daily and leisure activities
- (11) Birth order of patient and all sibs
- (12) Thumbnail sketch by interviewers of social situation

The draft of the Social Description schedule was presented to the meeting of collaborating investigators that took place in November 1967 in Geneva. They were asked to discuss and give their comments about the

usefulness of each item based on their experience of the relevance and availability of information for the items within the context of the socio-cultural characteristics of their centres. After this they had to "vote" on whether they thought each question should remain in the schedule or be deleted. Extensive discussions took place; it was interesting that here again the main issue was not so much which questions should be included but whether the questions should be pre-coded or open-ended. The disadvantage of pre-coding, as with all pre-coded questionnaires and rating scales whether used in national or international research, is that there is no way of checking the validity of the ratings or of making the coding specific for each centre. To avoid this danger, many questions were made open-ended. It was felt that this course might be the wisest to adopt while the schedule was being developed, and that pre-coded questions could be used at a later stage. With the new schedule that was used in the main part of the study, a fairly extensive set of instructions was produced. An example of a question and corresponding instruction was as follows:

Question A4a:

"Now I would like you to describe the place where you are (were) living. Is (was) it your own house, or do (did) you rent an apartment or some rooms or what?"

A4a Write description

Instructions

Try to get the patient to tell you a reasonably accurate, vivid description of where he is living. For example: "I live in my own house that we bought ten years ago and it is about 20 years old. It is a small wooden house and needs painting and we really don't have enough room now for my family as there are only two bedrooms in the house and all six of us have to use them for sleeping, etc."

It is realized that the Psychiatric History and Social Description schedules developed in the course of the Study were not perfect. However, they draw attention to the difficulties to be overcome before this underdeveloped area of investigation can be made fully productive.

5.3 Diagnostic Assessment Schedule

In Phase 1 of the IPSS, the psychiatrists were not asked to make a diagnosis but rather to say to which of three operationally defined groups they thought the patient belonged: definitely schizophrenic; definitely not schizophrenia; or possibly schizophrenia. They also had to state the reasons for their assignment. Where there was disagreement over the classification of a patient in simultaneous interviews, the reasons had to be recorded on a special form.

After the results had been reviewed, and because it had been decided to record a diagnosis for all patients in the main part of the study, it was necessary to develop a diagnostic assessment schedule. The schedule used in Phase 2 had four parts:

(1) The diagnostic formulation. After he had examined the patient and reviewed all the information about him available in the various schedules, the psychiatrist was required to make a diagnostic formulation in which he summarized the main findings in the psychiatric history, social description, physical examination, and present state examination. In addition he had to diagnose the patient, using the system with which he was most familiar and that he was using regularly in his centre.

(2) A checklist of diagnoses in which the categories for functional psychotic disorders from Section V of the International Classification of Diseases were listed. The psychiatrist was required to classify the diagnosis under one of the categories with the aid of the Glossary.

(3) A statement about the certainty with which the diagnosis had been made. If the psychiatrist was not certain about his diagnosis he was invited to state his reasons and give an alternative diagnosis.

(4) Three brief rating scales about the prognosis of the patient. These were related to the patient's condition in two years' time and subsequently, and concerned his needs for treatment, his condition if ideal treatment could be given, and his probable ability for self-support.

In reliability interviews each of the psychiatrists participating had to fill in a form so that the extent of their agreement could be assessed.

5.4 Physical and Neurological Examination Form

The Physical and Neurological Examination (PNE) was not among the most important areas of interest in the IPSS because, as specified by the screening procedure, patients showing any physical or neurological abnormalities possibly connected with their symptomatology had to be excluded. Hence it was only used to assess patients with abnormalities unrelated to their mental condition. The schedule could be completed by any medical officer.

The first version, used in Phase 1, was a very simple checklist of

5.5 Screening Forms

For the development and description of the Screening Forms, see Chapter 1.

5.6 Condition on Admission Form

It was recommended that the PSE be performed as soon as possible after the patient's admission to the psychiatric facility. This meant that in many cases the interview would be carried out before there had been an opportunity to complete the Psychiatric History schedule, especially where information for the latter had to be obtained from an informant or involved home visits. In order that the project psychiatrist should have at least some basic data about the patient, the Condition on Admission form was developed. This form had to be completed by the admitting psychiatrist or one attending the patient on the ward and could then be used when the Psychiatric History schedule had not been completed or was not available.

The information on the Condition on Admission form was designed to give some indication of the patient's pre-admission symptoms and behaviour and how long these had persisted. Also, as the state of the patient at the time of the PSE might be influenced by any medication he had received, especially since admission, the drugs or other treatment given to him were also recorded.

The version used in Phase 1 of the Study contained six items and a space for recording who was the informant in each case. The first question was, 'State why the patient was admitted to the facility'. It was hoped that the information elicited would shed light on difficulties in the home or work situation, for example, that made it impossible to keep the patient at home longer, or that made the doctor admit him. However, in the trial registration the answer usually given was 'Because of mental illness'. This question was therefore dropped from the revised form used in Phase 2. In any case, it could have considerably overlapped with the second question, which asked for a brief account of the patient's pre-admission behaviour. This question was not particularly successful either, because although the informant undoubtedly gave his replies in such form as "says he hears voices" or "complains that neighbours are saying he is a criminal", the interviewer recorded the corresponding psychiatric symptoms. For example, it is unlikely that the mother of one of the patients from a rural area would describe his behaviour as "social withdrawal based on delusions of reference and persecution". By such descriptions the psychiatrist initially looking after the patient would begin to bias the investigator before the latter commenced the PSE interview.

In both versions of the form a question was asked about who brought the patient to the psychiatric facility; it was felt that this person might be used as an informant for the patient's Psychiatric History and Social Description interviews and also might know the patient's whereabouts at the time of the followup study.

The final version of the Condition on Admission form had the following

six questions:

(1) Give brief description of pre-admission behaviour, e.g., spoke of suicide, violent, refused to eat, depressed, etc.

(2) How long had the patient been like this?

(3) Has patient had psychiatric care before? (ring)

1 Yes 2 No 3 Not known

(4) Was the patient willing to come to hospital? (ring)

1 Yes 2 No

(5) Who accompanied the patient to the facility?

(6) What ECT or drugs, and dosage, has the patient had since admission?

5.7 Training in the Use of Instruments

In an international study like the IPSS, two types of training had to be given: first in the organization and conduct of the study, second in the method of using the research instruments. How the study should be run was decided by joint discussion at the meetings of collaborating investigators. The main emphasis, however, was placed on the second type of training. Instruction in using the schedules was of two kinds — that given to interviewing psychiatrists at the beginning or during the course of the study and that given to other members of the research teams in the centres, including those who joined the staff of the FRCs while the study was in progress.

For the psychiatrists, training in the use of the PSE presented the greatest difficulty, since this schedule was not fully standardized either in its application or in its scoring. Hence the only way in which a psychiatrist could achieve reliability was by what amounted to a short apprenticeship to someone who was already proficient in applying and scoring the schedule.

Clinicians already familiar with the process of diagnostic examination needed to do about 20 PSE interviews in order to become thoroughly acquainted with the instrument, 10 of them under supervision. There is no substitute for the long process of undertaking varied interviews in the presence of a supervisor and then going through the ratings one by one in order to discuss discrepancies and to learn the intentions behind the way questions are worded, sections ordered, and rules and definitions specified. Although the PSE had gone through eight editions and been used in varied psychiatric situations around the world, it was plain that, no matter how clear the instructions might be, it was impossible to carry out the interview adequately simply from reading the schedule. It should be noted that a full set of instruction manuals, with illustrative audio and videotapes was prepared for training purposes.

As in all procedures involving clinical judgement, personal training remained essential.

Despite the administrative difficulties involved, two training seminars for the collaborating investigators were held in London in May and July 1967. By this time participating psychiatrists were acquainted with the schedule and several had translated it into their own languages. As a preliminary exercise, a videotaped interview was shown, with pauses between sections, during which members of the group discussed their ratings of each item in detail. This procedure was most effective for clarifying the meaning of questions and instructions and making explicit those points that could not be spelled out in detail in the schedule. Coding problems were also identified and dealt with. A second videotape was then used for a reliability exercise, again with detailed discussion on divergences in rating. The participants were then divided into four pairs in which each psychiatrist was asked to interview a patient "live" in the presence of the other and of an experienced user of the schedule. The four groups then held individual "port-mortems" and afterwards all groups came together to exchange views. Lastly, a third videotape was shown and the exercise repeated. This systematic and intensive introduction to the use of the PSE was well received by the participants and the reliability of the ratings made was satisfactory for this stage. The 26 interviews which were carried out in the course of Phase 1 served both for the assessment of the schedule and for further training of the investigators, particularly under "field" conditions.

Once some degree of uniformity had been achieved, it had to be maintained and if possible improved by the psychiatrists when working in their centres. They were therefore asked to carry out simultaneous interviews at regular intervals throughout the period when patients were being taken into the study, and to discuss their ratings, especially where they disagreed. The multiple interviews carried out during exchanges of visits and the ensuing discussions also promoted uniformity in rating and were useful for the assessment of reliability.

It had been intended to give one psychiatrist further training in the interview under supervision and then send him to each FRC in turn to sit in with the collaborating investigators when they were interviewing English-speaking patients and thus act as a referee for the whole study, with respect to the PSE. Since it was unfortunately not possible to arrange this, the question of how uniformly the interview was carried out in the centres cannot be fully answered. However, from the results (described in Chapter 8) of the reliability exercises carried out in the centres and of the multiple ratings of videotaped and filmed interviews done at exchanges of visits, it would appear that uniformity of interviewing was reasonably good.

The investigators were already familiar with many of the questions in the Psychiatric History schedule, which are those commonly asked in eliciting case histories. As the information required had to be obtained from various sources (patient, informants, case notes) the training required was different from that needed for use of the PSE. No supervised interviews took place, but instead the investigators were given detailed instructions on how to use

the schedule.

All research personnel, medically qualified or not, who were responsible for completing the Social Description schedule had to be trained in the centres by the collaborating investigators. This was especially necessary for those team members who spoke only the language used in their centres.

When new items were introduced into the schedules or when the scoring changed during the course of the study, the relevant instructions were discussed during the meetings of investigators and later circulated in the report of the meeting. The Instruction Manual for Phase 2 was prepared during a meeting of investigators held to discuss Phase 1, finalized in the light of the discussions, and modified only to allow for changes made in the schedules.

In some centres new psychiatrists or other staff had to be trained while the study was in progress. Psychiatrists joining the project had to do at least five simultaneous interviews with each of the other psychiatrists in their centre. By discussing their ratings it was possible to establish standards of rating among the new members similar to those among the original raters. In some cases the new member could be sent to another centre for training, so as to promote inter-centre reliability, but unfortunately this could not be done in every instance. However, during the course of the study the investigators had opportunities to complement their training.

5.7 Glossary of Psychopathological Terms

At a very early stage in the study it was pointed out that difficulties in communication between psychiatrists would stem largely from differences in their definitions of the terms and concepts used. In Phase 1 no diagnoses were made. As mentioned earlier, the psychiatrists were simply required to assign the patient to one of three groups: definitely schizophrenic, possibly schizophrenic, or definitely not schizophrenic. From an analysis of the results it was obvious that the same terms had different meanings for several of the investigators.

A content analysis of the description of the first 300 psychiatric patients produced a list of 47 terms that seemed to have been used to express very different things. These differences could be due to the psychiatrists having varying degrees of fluency in English, different theoretical backgrounds, or different conventions for the usage of terms.

In order to examine the differences between the "content" of these terms and to make the psychiatrists aware of them, a list of the 47 terms was sent to all the centres and the investigators were asked to define each of them. In response, some centres sent the definitions given by one or all individual collaborators, while others provided a 'centre definition'. In spite of the explicit request to define all terms, comments were received such as "clearly unnecessary to define" or "this term should be omitted", suggesting a conviction that the author's definitions or theoretical allegiance must be universally shared. Fortunately these comments were rather few, which augured well for agreement on a future list.

The analysis of the definitions sent to Headquarters showed both

unexpected agreement and also unusual differences in conceptual structures among the collaborators. Differences attributable to varying degrees of fluency in English were relatively few, considering that English was the first language in only two of the FRCs.

For a number of terms the differences in definition had nothing to do with language but were obviously due to different theoretical backgrounds, personal views, choice of different criteria for describing the term, and similar factors.

To ensure that terms would be used with the same meaning, a glossary of the terms frequently used by psychiatrists was produced. In developing the glossary, the more preferred definitions were chosen and modifications made so as to cover minor deviations from the most popular definition. In most cases the result was a definition that was reasonably similar to that of classical psychopathology. For example, in the case of "mannerisms" the final definition was: "odd and stylized voluntary movements, gestures or ways of expression"; for "elation": "affective state, characterized by feelings of profound extraordinary joy, cheerfulness, happiness and well being"; for "guilt feelings": "painful awareness of having committed an offence against one's moral code which calls for a punishment or retribution; pathologically excessive guilt about a minor or trivial offence often committed many years ago, about imaginary offences or a generally vague feeling of being guilty, sinful or evil without relating it to any specific event".

The glossary when complete was appended to the Instruction Manual and the collaborating investigators were instructed that, if they used the terms listed, they should then use them only in the sense defined.

5.8 Glossary of Diagnostic Terms

On the Diagnostic Assessment form the collaborating investigators were asked to make their own diagnoses in the terminology used in their everyday practice and then to classify them according to the International Classification of Diseases (8th Revision). Since some of the investigators were not familiar with the ICD, a glossary of diagnostic terms was appended to the Instruction Manual.

For the three digit categories in the ICD, a definition was first given for the whole group followed by the definition of each of the subgroups. An example is given below:

- 300.0 Neuroses - mental disorders in which the patient has considerable insight and does not confuse his subjective experience and fantasies with external reality.

- 300.4 Depressive neurosis - a state of depression which has recognizably ensued on and has been occasioned by depressing experiences; which does not include amongst its features delusions or hallucinations; which seldom leads to suicide; and in which there is preoccupation

with the psychic trauma that provoked the illness.

Insofar as the frequency and meticulousness with which the glossary was used, it appears that in some of the centres it was consulted whenever a diagnosis was made; whereas in others relatively little use was made of it.

TRANSLATION

An important preparation for the project was to translate the interview schedules into the seven non-English languages spoken by psychiatrists and their patients in the various centres: Chinese, Czech, Danish, Hindi, Russian, Spanish, and Yoruba. Translation of the documents into the local languages was done in the FRCs. Despite the difficulties described below, it was felt that the translations were satisfactory.

The translations had to satisfy at least two criteria. Since the wording of questions in the original English version provided the stimulus to the patient or informant to describe symptoms or history of disease, it was necessary to ensure that the stimuli given in all the centres would be as uniform as possible. At the same time, expressions and examples had to be chosen that would be appropriate in each particular centre. A basic problem is that the non-existence of an equivalent word may, and probably does, mean that the concept in question has little or no currency in the other culture, a fact that may play an important role in the formation of symptoms. The task of translating the schedules in an acceptable way was complicated by the variety of different idiolects, dialects, registers, styles, and modes of the languages used in the study (Catford, 1965).

The idiolect, i.e., the characteristics of a writer's style, might be a significant source of distortion in translation, in view of the fact that different writers composed the various instruments used. An idiolect could show itself, for example, in the excessive use of a particular term or turn of phrase. The translator's idiolect, particularly if the translation were very free, could change the entire meaning of an instrument used for psychiatric interviews.

The dialect used by the majority of patients in a centre was taken into account when translating the schedule. The translator obviously needed to be familiar with the peculiarities of this dialect, lest there be inconsistencies between the original version and the translation. This difficulty was compounded by the fact that a person capable of doing the translation might not speak the dialect used by many of the patients.

"Register" is a word that Catford uses to describe the specifics of a language relating to the social role of the speaker. The PSE schedule and some other schedules contain items with questions to be asked of a patient, i.e., in a "doctor-to-patient" register ("Did you feel people were against you", for example). There are also items containing titles of sections, rating scales, check-lists, and multiple choice questions for the use of professionals only, thus in a "colleague-to-colleague" register. This might have influenced the faithfulness of the translation in cases where the translator was not familiar with the specific jargon used by psychiatrists, psychologists, social workers, and others, both among themselves and when conducting an interview with a patient. It is interesting to note that most of the words that were reported as troublesome for translation appeared

in the "psychiatrist-to-psychiatrist" register:

<u>Danish:</u>	anxiety
<u>Yoruba:</u>	ritual, interest, depression, anxiety and tension mood, unreal, affect
<u>Chinese:</u>	anxiety, tension, depression, obsession
<u>Czech:</u>	delusions of reference

There were considerably fewer troublesome words in the "doctor-to-patient" register (e.g. thoughts: Chinese, Yoruba; worry: Chinese; future, unreal: Yoruba).

The style in which the original document is written can be an obstacle to a good translation that is difficult to overcome. The schedules used in this study make use of the consultative and casual styles (Joos, 1962) as well as the formal style. For example, compare the following two items from the PSE schedule:

7.01 Have you been very miserable, low-spirited or depressed during the last month?

14.14 Rapid and hurried succession of a number of actions, often without a logical sequence and without achieving result.

Spilka (1968) in her excellent article on translating a certain interviewing schedule divided the "components of translation" into (1) the text, (2) the translator, (3) the audience, (4) the source and target languages. She drew attention to the fact that both the co-text (i.e., the words immediately surrounding the word to be translated) and the context (i.e., the culture in which the interview is conducted) have to be taken into account. As far as the translator is concerned, she emphasizes how difficult it is to find a person who is sufficiently familiar with the linguistic usage of both parties involved in the use of instruments, namely the patients and research team members. Very often this difficulty can be overcome by employing several people to work together. Even so, there are difficulties that can be only partially solved. Spilka suggested various techniques and procedures intended to make the translator better able to translate highly technical material. In this study, specially trained translators were not available and this may have been an advantage because, as is pointed out later, non-professional translators may do better when translations of schedules and questionnaires are assessed for equivalence.

The question of audience is a particularly difficult one because of the wide socio-cultural and economic differences between the psychiatric patients who were to be interviewed with these schedules. In fact, although in some centres the groups of patients were fairly homogenous, it could be said that the variations among patients in this study were probably greater than in any group of patients who have been assessed with a set of standardized

research instruments.

The "source" language for all the research documents was English, which at least provided some uniformity. The "target" languages — seven in all — were, however, very different in their structure, availability of words, synonyms, and other characteristics. The problem of translation, which is difficult enough when only one target language is concerned, becomes considerably more complicated when instruments are needed in more than one target language because it becomes necessary to ensure equivalence among the versions in the target languages as well as with the version in the source language.

Various techniques can be used to assess the quality, equivalence, and usability of a translated version of a schedule or other document. Prince and Mombour (1967), for example, used a very thorough method to assess the equivalence of the English and French versions of Langner's 22-item scale of psychiatric impairment. They assigned 80 bilingual patients randomly to two groups and then applied the first half of the questionnaire in the source-language version to one group and the target-language version to the other. The second half of the questionnaire was given to each group in the other language. In 4 of the 22 items the frequency of positive answers differed significantly depending on which language was used, even though the two languages were similar in structure, syntax, and a number of words. It can well be imagined how significant the differences may become when the languages belong to less related families.

An even more striking example of differences that may be encountered when using schedules in translated versions was given by Ervin (1964). She showed that when bilingual subjects were given personality tests in their two languages they had apparently different personalities. Similar studies have shown that a variation in correlation from .53 to .9 can be expected when bilingual subjects are tested in their two languages.

Unfortunately, this method of assessing translations did not seem feasible in this study because, in addition to requiring bilingual individuals to be found in the various centres, they would also have had to satisfy the criteria of the psychotic screen, which would considerably reduce the number of subjects. It would also have been necessary to seek people bilingual in all the various combinations of languages, which is an impossible task. The number of psychotic bilingual Yoruba-Chinese or Hindi-Spanish speaking patients is probably infinitesimal, even if the whole population of the countries and not only those residing in the FRC catchment areas were screened.

A reasonably good way to ensure equivalence of translation is to have each item discussed intensively and at length by people speaking the source and target languages, some of them speaking both. A. Leighton et al. (1963) had to prepare seven versions of a questionnaire to be used in studies of psychiatric epidemiology before they felt that the translation from English to Yoruba was satisfactory.

In the IPSS, back-translation was used for assessing the quality of translation since it is generally considered to be a satisfactory method for this purpose. The schedule was first translated into the target language by one person; the translation was then given to another person who translated it back to English. The original and the back-translation were then com-

pared. This method has some disadvantages, but these can be overcome by giving them proper attention. One is that it is difficult to distinguish how much of the meaning has been changed in each of the two stages (i.e., how much in translation as opposed to back-translation), and whether to introduce the necessary amendments in the original schedule or in the translated version. Some African languages, for example, lack words describing colours (East, 1956), so that the original version has to be changed if colour words are used. In other cases the meaning of a word in the source language is either extremely narrow or else very broad and no direct one-word translation is possible.

Every language has such words and many of them are entering international usage without translation, e.g. "set", "bias". In psychiatric schedules, however, it is only rarely possible to avoid translation.

When dealing with behavioural sciences, and particularly psychiatry and psychology, it is of the utmost importance to ensure not only that the translation is faithful, but also that the words chosen in the target language carry the same connotation as in the source language. From the back-translation it is often extremely difficult to see whether this has been done. Dictionary equivalents are used very often, particularly if skilled translators work on the translations and back-translations, so that the same term will appear in the back-translation as in the original document in spite of the fact that the equivalent in the target language sounds stilted, has a different connotation, or is completely out of place. Non-professional translators usually do better, simply because they do not have the professional translator's skill and his ability in the use of dictionaries, glossaries, and other aids.

Several translations and back-translations may be necessary in the process of producing a document for use with patients or other informants, particularly if the matters assessed are so dependent on the interpretation of words by the interviewer and by the interviewee as is the case with psychological investigations.

There are a number of rules that have to be kept in mind when preparing a schedule to be translated for international use. Werner and Campbell (1971) have listed some of them: the sentences must be simple and passive voice avoided when possible; subjunctives should be avoided; metaphors and colloquialisms should not be used; nouns should be repeated rather than replaced by pronouns; hypothetical phrasing should be avoided.

The analysis of a schedule that is to be translated, from the point of view of feasibility of translation, can avoid lengthy and expensive changes and amendments at a later stage.

Longacre (1958) showed that words are mostly translated poorly and equivocally when they are alone; they are translated better when they are in a sentence and better still in a paragraph. This argues against adjectival checking lists, against the application of single word polar scales, and particularly against the repetition of specific words in various items, because of the possibility that an imperfection will be multiplied. Whenever possible, multiple synonyms should be employed rather than the same single term.

The normal translation—back-translation procedure with 3 steps (source-

target-source) may be insufficient to guarantee a reasonably satisfactory translation. Werner and Campbell's (1971) remedy for this was to have the three steps performed several times ("iterative back-translations") and to approach equivalence in this way. Brislin (1970) preferred a different method and introduced an additional step so that a 4-step procedure results: source; target; target check; source—back-translation. The "target check" should include a search for errors of meaning. Brislin suggested that the "meaning error" can be taken as a unit for measuring the quality of translation and defined it as the number of errors that would lead to differences of meaning according to the judgement of monolingual raters.

What can happen if even one of the above rules is not observed may be illustrated by an example from the back-translations of the PSE schedule. An item in the section on delusions of control reads:

...Does some other force than yourself make you do, feel or say things that you do not intend? As though you were an automaton, robot (zombie), marionette, puppet, without a will of your own?

The words "zombie", "automaton", and "robot" do not lend themselves readily to translation. They are rather specific and though probably well understood by a middle-class European might not be grasped by some patients included in the study. When translated and then back-translated, the second part of the item became as follows:

Back-translation from Spanish:	Do you feel like an automaton, or a mechanical doll (living death, marionette, puppet)?
Back-translation from Yoruba:	As if you were an image without your own will (fairies, image and others)?
Back-translation from Danish:	As if you were an automaton or a robot without a will of your own?
Back-translation from Hindi:	As though you are an automatic instrument or a puppet which does not have its own will (zombie, marionette, puppet, other)?
Back-translation from Chinese:	As though you were an automatic machine or a robot without a will of your own (some supernatural power that revives the corpse, marionette, puppet, others)?
Back-translation from Czech:	As if you would be some automaton, robot-machines (a plaything, a doll)?

Fortunately, there was sufficient range of alternative terms in this item so that it was probable that the meaning was clear when the whole item was used. This is a general rule, which is unfortunately often neglected in presenting schedules for transcultural use: the more words there are in

an item the less the chance that it will be misinterpreted.

An item that was to be rated after the delusional section had been completed read in English as follows:

- 12.1 Do you think you deserve these experiences?
(Elaborate in subject's own terms)

The back-translations were:

from Chinese:

Do you think you deserve these experiences?

from Hindi:

Do you think that you are fit to have such experiences?
(Elaborate in subject's own terms)

from Spanish:

Do you think you deserve what is happening to you?
(Elaborate in patient's own words)

from Yoruba:

Do you think you deserve these experiences you are undergoing?
(Explain fully what patient says)

from Danish:

Do you think you deserve such treatment (voices, visions, etc.)?

from Russian:

You think you have deserved these feelings (experiences)?
Record the patient's formulation.

There are two serious differences between the original and the back-translated versions. The first is in the back-translation from Hindi, where the word "fit" replaced the original "deserved". It is not difficult to see how a two-step transformation of meaning could have arisen. "Do you deserve such experiences?" could be interpreted by a person whose mother tongue is not English as "do you feel you were not bad enough" -- if they were so unpleasant as to be a punishment. If the first interpretation is taken, the back-translator has to convey in English the idea of whether the subject merits the experiences, which is done by using the expression "Are you fit to have the experiences". Obviously words such as "deserve" or "merit" are open to a number of interpretations and should be avoided. This item was left out from the revised version of the schedule.

The replacement of "experience" by "treatment" in Danish has been corrected by adding examples ("voices, visions"), but only partially. It should be mentioned that the examples do not appear in the original test. This shows another characteristic feature of translation: the non-professional translators add words to clarify sentences that may have been unclear in the original, a possibly useful technique if consistently employed, but

dangerous in some instances where the examples can mislead the interviewers.

The second difference in this example concerns the instruction: the word "elaborate" should not have been used in a schedule meant for translation because it is not clear to what "elaborate" refers. For example, to the uninformed this instruction may mean either to ask this question in words familiar to the patient or to record the answer in the patient's words.

When an item is formulated in a simple way, with sufficient redundancy of terms and with simple words, the back-translations are much more like the original:

Original

11.4 Have things looked dark, or grey, or colourless?

Back-translations from:

Chinese:

Do things look dark, grey or colourless?

Czech:

Do things appear grey or colourless?

Danish:

Have things looked dark, grey or colourless?

Hindi:

Do objects look deep coloured or grey, without colour?

Russian:

Do things look dark, grey or colourless?

Spanish:

Have things seemed to you to be grey, or dark or without colour?

Yoruba:

Are things looking dark, or grey or colourless?

A thorough and complete assessment of the results of translation of documents into the seven different languages of the project had not been completed by the writing of this report. There is, however, ample indication that the translated versions of the documents have been satisfactory and that equivalence, both formal (of form and content) and dynamic ("concerned with the receptor's response"), was achieved (Nida, 1964):

(1) The "traget-check", which some authors (e.g., Brislin) consider very important, has been carried out frequently in all of the FRCs by psychiatrists and other staff using the schedules.

(2) Most of the members of the research team had at least some knowledge

of English; a considerable number were fluent in both languages.

(3) In the statistical analyses that were carried out in order to compose the Units of Analysis (see Chapter 7), very similar patterns of correlations were found in the various Centres. This would not have been the case if the schedules were conveying different meanings in the various languages.

(4) The psychiatrists, and through them the other members of staff, have been instructed about the meaning of the items in the schedules and the manner of using them.

In a number of instances the psychiatrists conducted the interview in a local language and made the ratings on an English version of the schedule. The interviews on films and videotapes used for reliability exercises were both carried out and rated in English, and good agreement was obtained. This means that the raters understood the same questions in the same way.

It is clear that a careful and detailed analysis of the equivalence of the schedules in the various languages will have to be carried out before the translated versions can be released for general use. At the present stage of the work it appears that the schedules used in the different centres have been of satisfactory linguistic equivalence.

UNITS OF ANALYSIS

For each of the 1,202 patients in the IPSS, 360 items were rated in each Present State Examination (PSE). This large body of information had to be condensed to a manageable amount. Clinicians and scientists think in terms of symptoms, syndromes, and diagnosis rather than in terms of answers to individual questions that are only one means for deciding whether a symptom is present or not. Terms that are easily understood and handled must be used, both in interim analyses that aim at confirming or discarding a working hypothesis and in reporting results to other scientists. Moreover, without condensation the enormous volume of data could not be easily managed by computers of an ordinary size; even if it could, it is almost unthinkable that someone could read and interpret the output within a useful time period.

Another very important reason for seeking ways of condensing information and finding out which items can be merged with others with a minimal loss of information was the need to arrive ultimately at a shorter schedule. A main goal of the project was to produce instruments that could be used in epidemiological studies of defined populations. The existing PSE took from 45 to 90 minutes to complete, apart from the time spent on the Psychiatric History (PH) and Social Description (SD) interviews.

7.1 Relative Advantages of Methods of Grouping Items

Data can be condensed in 3 ways:

(a) Items can be grouped or selected on the basis of clinical judgement to form symptoms or syndromes.

(b) Mathematical techniques, such as factor analysis, can be used.

(c) Methods (a) and (b) can be combined and groups composed on the basis of both clinical judgement and the results of statistical analyses.

(a) The grouping of items on the basis of clinical judgement alone may have the advantage of producing syndromes that are clinically meaningful. Clinical judgement can, of course, err from a statistical point of view: items may be grouped together simply because each is directly related to another common, underlying, and diagnostically significant factor. Any grouping suggested on a purely clinical basis would need to demonstrate its value empirically, for example, by leading to a useful clinical classification.

Psychiatrists are rarely in complete agreement about the grouping of items. Thus, when the collaborating investigators were asked to make suggestions for grouping, several lists were sent in that reflected divergent views. The first difference of opinion to emerge was whether items from

the behavioural sections of the PSE should be grouped with items from the non-behavioural parts and whether items from below the cut-off points should be included. There was further disagreement about whether to include an item in a group if it had either 1 and 2 ringed, or only if 2 were ringed.

The use of clinical judgement as a sole criterion for the composition of the groups also means that the various factors known to influence judgements can come into play, such as (1) "logical error", when items are grouped together because it is "logical" that they should be together, e.g., items on delusions of persecution and social isolation; (2) the "hypothesis error", where items are grouped together because there is a hypothesis that they are an expression of the same underlying phenomenon (e.g., items on homosexual leanings and delusions of persecution, which are considered to be causally connected); (3) the "error of association", when items that are often found together will be put together, although they do not represent the same phenomenon (e.g., insomnia and loss of weight); and (4) "personal error", where the clinician's personal idiosyncrasy plays a role in putting items into groups. Nevertheless, investigators often group items on the basis of clinical judgement because such groups can be established quickly and result in a significant reduction of the number of items that have to be dealt with. In this way the items can be translated into clinically meaningful symptoms. The CATEGO procedure described in Chapter 11 is based upon symptoms generated in this manner. The same kind of clinical skills that produced the PSE were used to group the PSE items.

(b) Purely mathematical methods of grouping items are free from the influence of the factors referred to above but may have the disadvantage of producing groups that are clinically meaningless or that can only be explained with the aid of complicated assumptions.

When correlational methods of grouping items are used, the nature of the mathematical relationship to be discussed has to be carefully considered. Linear relationships are often assumed to obtain, but in cases of mental disorder this need not be so: for example, minimal motor activity may be associated with maximum depression in retarded depression; and maximum motor activity will be associated with maximum depression in agitated depression; minimal depression will correlate with average psychomotor activity. In this case the relationship is quadratic.

Although it is possible for a mathematical technique of grouping items to assign weights to particular items, it is very difficult to propose a weighting system that would consider all the variables a clinician uses in assessing the significance of an item.

Nevertheless, mathematical procedures have the advantage of searching for significant groupings of items without making any of the assumptions referred to above in the discussion of the use of clinical judgement, and may in fact produce combinations that either suggest new ways of grouping patients or help to validate the present methods. Carried one stage further, this line of reasoning leads to classification by mathematical techniques, discussed in Chapter 12. Again, the final value of any groupings thus produced must be assessed empirically.

(c) It was therefore decided to use a combination of clinical judgement and mathematical techniques in grouping items in the IPSS. The items were

first put together on the basis of clinical judgement. Each of these groups of items was tested using data obtained in the study. If it was found that the statistical correlations between the items in the units did not confirm that the items belonged together, the items were regrouped and the statistical tests repeated. This sequence — clinical grouping - statistical test - new clinical grouping — was repeated until a unit was obtained that was clinically meaningful and empirically tested for association among its components. Such a unit was termed a Unit of Analysis (UA).

A comparison among units obtained by the three methods described above (i.e., clinical only, mathematical only, combination clinical-mathematical) will illustrate the differences among them.

The grouping of items described in Chapter 11 (CATEGO) is based on clinical judgement. An independent factor analysis (FA) carried out by Fleiss et al. (1971) on data obtained in the U.S. - U.K. Diagnostic Project on a schedule similar to the PSE is an example of a mathematical method of grouping items. The groupings that result from these two methods are compared below with the results of the UA method in terms of one UA, "Delusions of Persecution".

<u>PSE Number and Title</u>	FA	UA*	CATEGO
12.21 Dropping hints with special meaning	x		
12.25 People not what they appear to be, perhaps disguised	x		
12.26 Things arranged to convey special meaning	x		
12.30 Suspicious of people's intentions			x
12.31 People want to harm	x	x	x
12.32 Force trying to act, harm		x	x
12.33 Followed around, spied upon	x	x	x
12.34 Someone trying to poison	x	x	x
12.35 Particular person trying to harm		x	x
12.36 Harmful organization - Mafia, etc.		x	x
12.37 Spreading gossip	x		
12.39 Machines act to detriment		x	x
12.40 Singled out for persecution	x	x	x
12.41 Other delusions of persecution	x		x
12.42 Intention to destroy patient			x

* The UA also contained the item 12.38 - Machines follow patient about, but this has been left out here because it did not appear in the other analyses.

It will be seen that the purely mathematical method of analysis leads to the inclusion of four items (12.21, 12.25, 12.26 and 12.37) which clinically do not belong to "Delusions of Persecution", while in CATEGO analyses two items are included that do not appear in the FA or UA groups. However, in the CATEGO analyses this group of items is divided into two subgroups:

Delusions of persecution by people 12.30, 12.31, 12.35, 12.36, 12.40,
12.42

Other delusions of persecution 12.32, 12.33, 12.34, 12.39, 12.41

which are then combined into the one syndrome "Delusions of Persecution".

Thus the method used to produce UAs and the clinical method used in CATEGO arrive at almost identical solutions, while the purely mathematical method incorporates some items that appear to be delusions of reference rather than persecution and leaves out some items that, clinically, are obviously delusions of persecution.

7.2 Units of Analysis

The feasibility of grouping items into larger units on the basis of both clinical judgement and statistical verification was first tested on the items from the PSE schedule. Later, a similar process was developed for the PH and SD schedules. It might be of interest to describe here the procedure for arriving at the composition of the UAs in more detail because these have subsequently been used in a large number of analyses.

This procedure had several stages:

Stage 1. All participating psychiatrists were requested to group the items in the PSE schedule into clinically meaningful groups corresponding to symptoms. A list of this kind was also produced at Headquarters. Three of the FRCs produced such lists; others adopted the Headquarters' list with only slight modifications.

Stage 2. The four lists were studied in detail and a compound list was produced that took into account most of the suggestions from the proposed lists. It was encouraging that this could be done because it meant that the psychiatrists producing lists or suggesting changes in them were in good agreement about which items compose symptoms, despite the fact that they were trained in different schools of psychiatry and working in different countries.

Stage 3. This list was then assessed statistically, using the collected data, in two stages. In the first phase the UAs were tested using schedules concerning the first 700 patients, and in the second phase data for all 1,202 patients were used. Only one schedule per patient was used, the observers' schedules in simultaneous interviews and the second of two consecutive interviews being excluded. All the "other abnormality" and "overall rating" items were excluded from these analyses and treated separately, thus

leaving approximately 300 items that were grouped in 129 UAs.*

Indices of Association (IA) of ratings between each possible pair of items were computed for each UA. The formulae used for any given pair of items were:

$$IA(1) \text{ for items 1 and 2} = \frac{\text{number of schedules in which both 1 and 2 are positive (11) or negative (00)}}{\text{total number of schedules (00 + 10 + 01 + 11)}}$$

$$IA(2) \text{ for items 1 and 2} = \frac{\text{number of schedules where both 1 and 2 positive (11)}}{\text{number of schedules where item 1 positive (10 + 11)}}$$

$$IA(3) \text{ for items 1 and 2} = \frac{\text{number of schedules where both 1 and 2 positive (11)}}{\text{number of schedules where item 2 positive (01 + 11)}}$$

The values of each index could vary from 0 (no association) to 1 (complete association). The statistical significance of association was tested by computing the chi square and a Kappa test for each pair (Cohen, 1968). All three indices were examined at each stage of the analysis.

The decision as to when to consider an item positive for the purposes of calculating the index of association was based on clinical judgement and on the frequency of positive ratings, either "1", meaning the symptom was mildly present, or "2", meaning it was markedly or continuously present. If a large number of patients had an item rated positive, the power of the item was increased by taking only rating 2 to make the unit positive.

The rules for compiling and testing the UAs may be summarized as follows:

(1) Items were grouped on the basis of clinical judgement.

(2) Using all data (first from 700, then from 1,202 patients), indices of association were computed, and chi square and Kappa were calculated.

(3) Where a 5% level of statistical significance of association as measured by chi square and Kappa was not reached by an item in its first unit, the item was tried in another UA with a view to finding one where it would have a significant association with other items. In some cases items were tried in as many as eight different UAs before an optimum index was obtained.

(4) Where an item did not reach a satisfactory level of significance wherever it was tried, and there was no clear clinical reason for this, it was made into a single-item unit. There were in all 78 such UAs.

* For some analyses in this volume, 124 UAs were used, omitting biological treatment, unwillingness to cooperate, inadequate description, environmental circumstances causing cooperation difficulties, and speech impediments causing cooperation difficulties.

(5) Step (3) was repeated:

- (a) using the data for the patients in each FRC separately;
- (b) using the data for each rater separately;
- (c) using the data for three diagnostic groups separately and testing them across raters and across centres.

These groups were schizophrenia, depressive psychosis, and mania.

(6) In addition to the empirical composition and testing of UAs by the process described above, three more analyses were undertaken. They were:

- (a) measures of the correlation of all possible combinations of items and UAs;
- (b) measures of the indices of association between all possible pairs of items in the schedules;
- (c) measures of the correlation between all possible pairs of UAs.

It should be emphasized that this operational production of the UAs was performed twice over, first on the preliminary data from 700 patients and then on the total data from 1,202 patients. This process was expensive in terms of both the computer time and the man-hours required to examine the results and re-position the items. Nevertheless, it was regarded as essential because the UAs were to be used in many subsequent data analyses.

In the final list of UAs there were 46 that contained more than one item. Significance of association for the 367 pairs of items in these UAs was tested. 17 pairs were found not to be significantly correlated, 10 others were significant at the 5% level and the remaining were significant at the 1% level. It was found that in several UAs there was a group of items correlating significantly with each other, but that another item might correlate well with all but one. The reason for this could often be found -- for example, the items represented events occurring at different times, such as "anxiety observed in the interview" and "reported anxiety".

Since the UAs represent symptomatology, combinations of items might be chosen that, if positive, would guarantee the presence of the symptom but that could clearly not correlate. An example of such a UA is derealization, which includes "things look grey, colourless" and "things look colourful". These did not correlate significantly with each other but both of them correlated significantly with "feels surroundings have changed".

7.3 An Illustration of the Procedure

To illustrate the procedure described above, the successive "runs" (i.e., re-groupings) for the UA "overactivity and agitation" are shown below. The first series of calculations was based on data for 700 patients,

and combined data for all diagnoses and all centres. In Tables 7.1-7.4 the second formula for the index of association was used. For the final calculations based on 1,202 patients, the numbers of positive ratings are given.

Table 7.1. OVERACTIVITY AND AGITATION (700 SCHEDULES)

Item number in schedule and description	Item number in schedules							
	8.3	14.11	14.12	14.13	14.15	14.17	14.18	14.37
8.3 Felt restless	1.000	.543	.472	.542	.429	.444	.440	.522
14.11 Can't sit still		1.000	.764	.739	.857	.844	.864	.696
14.12 Paces room			1.000	.557	.857	.778	.692	.652
14.13 Fidgeting, hand wringing				1.000	.643	.611	.538	.652
14.15 Violent excitement					1.000	.444	.231	.217
14.17 Shouting, screaming						1.000	.231	.348
14.18 Singing							1.000	.174
14.37 Grips chair, hugs self, clenches fists								1.000

It seemed that several items from this (clinically composed) group fitted in very well with each other, namely 14.11-14.13, 14.15, and 14.17, whereas 8.3, 14.18, and 14.37 did not give such good indices of association with some of the other items.

It may happen that a group of items put together on the basis of clinical judgement may show relatively low associations between its components. Such low correlations can in some cases be explained by the structure of the items. In other cases they are an indication that the items do not "belong" together, i.e., that they are not facets of the same symptoms. In the above example it might be expected that 14.15, violent excitement, would associate well with 14.17, shouting and screaming, whereas 14.15, violent excitement, might not associate as well with 14.37, grips chair, etc., but better with

14.11 and 14.12, which indicate more active behaviour. Nevertheless, 14.37 in itself is behaviour indicative of agitation. The fact that item 14.15, violent excitement, has a relatively low correlation with the other items is understandable, since patients who may have had this symptom in other settings are less likely to show it in the interview situation.

From Table 7.1 it is clear that the non-fitting items in this unit were 8.3, felt fidgety and restless, and 14.37, grips chair, etc. The former was dealt with as follows:

Item 8.3. The frequency of positive ratings for item 8.3 was very high among the first 700 patients. Where there was a positive rating, in 80% of the schedules it indicated mild restlessness, reported to have occurred in the course of the last month. It was reasoned that if the rating "2" only (continuous or severe) on item 8.3 were taken as making the item positive a better agreement between 8.3 and the other items, all of which relate to observed behaviour, could be expected. The association of this item, where it had the rating 2, was therefore tested with the following result:

Table 7.2. OVERACTIVITY AND AGITATION (700 SCHEDULES)

Item number in schedule and description	Item number in schedule							
	8.3	14.11	14.12	14.13	14.15	14.17	14.18	14.37
8.3 Felt fidgety and restless; (rating 2 only taken as positive)	1.000	.420	.211	.253	.214	.333	.240	.174

Since the associations were lower than before, a new "home" was sought for item 8.3. It was now tried with item 14.11, taking a rating of 2 only as positive, and also with item 15.01, the emphasis of the UA being changed from "overactivity" to "signs of tension" (Table 7.3). This did not give any better results. Item 8.3 was next tried with other items of reported tension, 8.1 and 8.2 (Table 7.4). To save time, the association of these with observed tension were also computed at the same "run". The associations between 8.1, 8.2, 8.3, and 15.01 were now sufficiently high and they were therefore put together in the UA "Tension".

Similar procedures were carried out with other items that did not fit in with those in the UA as originally proposed. The result of these processes of hypothesising, grouping, testing, and re-grouping done with all items in the PSE was a list of UAs, version 1, containing 129 UAs. The UAs were used in a number of interim analyses (e.g., assessment of reliability, symptom profiles, etc.) and in testing methods for data analysis (e.g., cluster analysis).

Table 7.3. TRIAL OF UA "TENSION" (700 SCHEDULES)

Item number in schedule and description	Item number in schedule			
	8.3	14.11	14.12	15.01
8.3 Felt restless (rating 2 only taken as positive)	1.000	.225	.211	.214
14.11 Can't sit still (rating 2 only taken as positive)		1.000	.764	.322
14.12 Pacing room			1.000	.281
15.01 Facial expression tense				1.000

Table 7.4. PROVISIONAL UA "TENSION" (700 SCHEDULES)

Item number in schedule and description	Item number in schedule					
	8.1	8.2	8.3	14.11	14.12	15.01
8.1 Unable to relax (reported)	1.000	.854	.713	.412	.374	.542
8.2 Muscles tense (reported)		1.000	.565	.309	.341	.393
8.3 Felt fidgety and restless			1.000	.543	.472	.560
14.11 Can't sit still (observed)				1.000	.764	.322
14.12 Pacing room (observed)					1.000	.281
15.01 Tense facial expression (observed)						1.000

As soon as all the schedules were received and processed at Headquarters the second series of testing of the UAs was started, using this time the total material from all 1202 patients.

The results of testing a provisional UA "Overactivity and Agitation" are shown in Tables 7.5-7.7. In these tables, whichever index of association was higher (2) or (3) was used. This procedure had two advantages, one immediate — it showed how often items appeared together, the other for later use — it showed which item was the better indicator of the presence of a symptom.

The associations shown in Table 7.5 were somewhat similar to those obtained for the first 700 patients, and all were satisfactory. The correlation between "violent excitement" and "shouting and screaming" was based on a small number of cases, but, as pointed out above, such behaviour during the interview is unlikely. However, in view of the fact that the final UAs were to be used for many subsequent analyses, further steps were taken to check the various sources of variation. The UAs were first analysed by diag-

Table 7.5 TRIAL OF ASSOCIATION OF ITEMS IN PROVISIONAL
UA "OVERACTIVITY AND AGITATION"
(1202 SCHEDULES)*

Item number in schedule and description	Item number in schedule				
	14.11	14.12	14.13	14.15	14.17
14.11 Can't sit still	1.000	.778 (293)	.682 (299)	.833 (254)	.913 (253)
14.12 Pacing room		1.000	.519 (262)	.833 (192)	.826 (193)
14.13 Fidgeting, etc			1.000	.611 (162)	.652 (163)
14.15 Violent excitement				1.000	.556 (31)
14.17 Shouting, screaming					1.000

* In Tables 7.5 - 7.7, the figures in parentheses refer to the number of positive ratings in both items.

nosis, beginning with schizophrenia. The associations (Table 7.6) were at least as good as for the whole group of patients. The next diagnostic group tested was depressive psychosis, with data from 99 patients from all FRCs. The indices of association were satisfactory. Although the calculations were in this case necessarily based on smaller numbers, there were at least 10 ratings for each calculation except one. Of the 10 indices concerned, 7 were 1.000 and the others .818, .636, and .471. Finally, the UAs were tested for the diagnostic group mania, with data from 66 patients from all centres. Here 5 indices were 1.000 and the others ranged through .889, .867, .500, and .467.

The analyses were also done for all the patients of each rater, irrespective of diagnosis. For those raters who had more than 20 patients in the same diagnostic group, additional runs were done. For example Rater 21 diagnosed 25 patients as schizophrenic. The indices of association for the UA "Overactivity and Agitation" for this group of patients were 1.000 in 8 cases, .900 and .667 for the remaining indices.

There were, however, numerous differences between the raters. In addition, when the units were tested using all patients from the same FRC (either by diagnostic group or all together) certain differences became apparent that made it necessary to re-order and re-group the items so as to achieve

Table 7.6 TRIAL OF ASSOCIATION OF ITEMS IN PROVISIONAL UA "OVERACTIVITY AND AGITATION" FOR SCHIZOPHRENIC PATIENTS (811 PATIENTS FROM ALL COUNTRIES)

Item number in schedule and description	Item number in schedule				
	14.11	14.12	14.13	14.15	14.17
14.11 Can't sit still	1.000	.758 (224)	.713 (218)	.813 (190)	.875 (189)
14.12 Pacing room		1.000	.574 (199)	.813 (156)	.875 (155)
14.13 Fidgeting, hand wringing			1.000	.625 (114)	.688 (113)
14.15 Violent excitement				1.000	.563 (23)
14.17 Shouting, screaming					1.000

an optimal balance of association between the items in each UA; this would make them most suitable for varying types of analysis. Although this exercise was time-consuming, both for scrutinizing the results and re-grouping items and also because, in all, as many as 20 runs were required for each UA, it was nevertheless felt to be essential for the success of future work.

An interesting side benefit of these analyses was the finding that raters could be clearly distinguished by the indices of association based on the schedules they rated. This finding can be used for analyses of the "personal formulae" of the raters and in the analyses currently in progress, which will make it possible to correct for personal bias and rating habits.

The final composition of the UA "Overactivity" after several further re-groupings (there were as many as 10 re-groupings in some UAs) was as shown in Table 7.7.

The total frequencies of positive ratings upon which the indices of association are based are fairly low, but such low frequencies of positive ratings of the items in the PSE schedule are characteristic in this study.

Table 7.7 FINAL COMPOSITION OF UA "OVERACTIVITY"
(1202 SCHEDULES)

Item number in schedule and description	Item number in schedule						
	14.11	14.12	14.13	14.14	14.15	14.17	14.37
14.11 Can't sit still	1.000	.725 (64)	.697 (90)	.846 (82)	.611 (87)	.565 (90)	.800 (81)
14.12 Pacing room		1.000	.485 (68)	.769 (54)	.500 (60)	.565 (61)	.600 (53)
14.13 Fidgeting, pacing room			1.000	.769 (36)	.278 (46)	.217 (51)	.600 (35)
14.14 Psychomotor pressure				1.000	.308 (27)	.308 (32)	.400 (16)
14.15 Violent excite- ment					1.000	.556 (31)	.400 (21)
14.17 Shouting screaming						1.000	.200 (27)
14.37 Grips chair, etc.							1.000

For example, the number of positive ratings of some of the items most commonly associated with schizophrenia are as follows (the maximum number of positive ratings on any item is 1202):

Presence of auditory hallucinations - 1 question	398		
Delusions of control - 3 questions	242	138	136
Delusions of persecution - 5 questions	452	424	206
	304	181	
Delusions of reference - 7 questions	491	449	212
	260	308	260
	147		

Such low numbers of positive ratings are, however, to be expected, since each patient has a limited range of symptoms.

It will be noted that some of the indices of association in Table 7.7 are lower than those in Table 7.1 involving data from 700 patients. The differences are due to the different numbers and different diagnostic and centre distributions of patients.

7.4 Further Statistical Assessment

As stated above in the rules for composing the UAs, three more analyses were undertaken to assess them:

(1) Measures of the correlations of all possible combinations of items and UAs were calculated. The correlation between UAs and items composing them were in general very high. There were, of course, some frequent associations of items with UAs other than those in which they were placed; most of these were explicable and often necessary associations, for example, items on delusions with the item on insight which was a separate unit in itself.

(2) The indices of association between all possible pairs of items in the schedule were determined. The highest associations between items were found between items placed in the same UA. Within-unit associations of items were for the most part higher than between-unit associations; where this was not the case, the connexions were predictable and justified in the light of clinical experience. It was to be expected that there would be connexions between items and UAs because this study was largely restricted to patients with functional psychoses who would therefore have a number of common characteristics of such psychoses.

(3) The correlations between all possible pairs of UAs were computed. These produced further confirmation of the clinical meaningfulness of these units. UAs were correlated in clusters that corresponded to syndromes of known psychopathological description. The results of this analysis were useful later when groups of UAs were produced (see section 7.6).

The final list contained 129 UAs, since in addition to the 124 UAs descriptive of symptoms 5 were developed describing circumstances that might have influenced the quality of information obtained in the interview and hence the importance to be attached to the results. They were as follows:

- | | |
|--|------------|
| (1) Patient under biological treatment | (3 items) |
| (2) Patient unwilling to cooperate | (5 items) |
| (3) Patient's description of problem inadequate, speaks vaguely or guardedly | (10 items) |
| (4) Cooperation difficulties, circumstances related | (4 items) |
| (5) Speech impediments | (2 items) |

A list of the UAs in their final version is given in Table 7.8.

The composition of the UAs produced by the method described above was now regarded as fixed in so far as the IPSS was concerned. However, just as an individual item may be regarded as positive, negative, or indeterminate on the basis of the code number ringed, so the same types of condition have to be defined for the UAs. This was done in different ways, according to the purpose for which the units were to be used. The easiest case was to make a simple dichotomy of the positive-negative (1-0) type. In this case a UA was considered positive if any of the items composing it was positive, i.e. rated 1 or 2; all other ratings were taken as making it negative.

In a second method of scoring the UA was counted positive (1) if any one of the constituent items was positive; it was counted negative (0) if all items were rated either 0 or not applicable, and in all other cases as indeterminate (9).

In some analyses a scaling system was adopted (e.g., see clustering techniques, Chapter 12).

Finally, provision was made for four ratings to be assigned to a UA: it was counted positive if any one of the constituent items was rated positive; negative if all items were rated 0 or not applicable; indeterminate if there were no positive ratings and any one of the items was not scored; and uncertain if all items were rated questionable. This method has been used for the majority of analyses in the study.

Thus, in summary, this approach is based upon both clinical and statistical methods. One psychiatrist was responsible for the clinical decisions, which were numerous and could not be specified in detail, although the principles used had been laid down. There is no certainty that, if another clinician had been involved, the UAs would have been composed in precisely this way. However, the clinical decisions made in this approach have been tested using statistical methods, and furthermore comparison with a purely clinical method, as used in the CATEGO program described in Chapter 11, indicates considerable agreement on the more important symptoms.

7.6 Groups of Units

The 360 items of the schedule were thus condensed into 129 UAs. The reasons for the condensation of items mentioned on page 94 apply for certain analyses to the UAs also, and a further condensation was therefore undertaken.

There are at least two possible methods of combining the UAs. UAs that appear together in clinical conditions can be grouped according to symptom patterns (syndromes). Thus, "delusions of control" and "auditory hallucinations in the third person" would be included under the "nuclear" syndrome. On the other hand, psychopathology can be broken down into major areas of dysfunction, and the units making up such areas can be grouped together. Using this method, "delusions of control" would be included under "delusions", and "auditory hallucinations in the third person" would be included under "hallucinations". For purposes of description, this method of grouping is called "groups of units".

Both approaches can be used quite profitably for the description and comparisons of groups of patients, for the assessment of reliability, and for various other analyses. Extensive examples of the use of Groups of Units (GUAs) are given in Chapter 10.

A list of groups of UAs, with their constituent UAs and the PSE items contained in each, is shown in Table 7.8.

TABLE 7.8 COMPOSITION OF GROUPS OF UNITS OF ANALYSIS

Groups	Units of Analysis	Items (indicated by PSE number)
1. Quantitative psychomotor disorder	1. Overactivity 2. Retardation 3. Stupor 14. Repetitive movements	14.11*, 14.12*, 14.13*, 14.14*, 14.15, 14.17, 14.37* 6.1*, 6.3*, 14.05*, 14.06*, 14.07*, 14.08*, 14.09* 14.10 14.30
2. Quantitative psychomotor disorder	4. Negativism 5. Compliance 7. Stereotypes 9. Grimacing 10. Posturing 11. Mannerisms 12. Hallucinatory behaviour 13. Waxy flexibility	14.41, 14.42 14.33, 14.34 14.29 14.36 14.27 14.28 14.40 14.32
3. Quantitative disorder of form of thinking (and speech)	15. Flight of ideas 16. Pressure of speech 18. Mutism 19. Restricted speech 124. Distractibility	16.33 16.13* 16.01 16.11 14.39
4. Qualitative disorder of form of thinking (and speech)	20. Neologisms 21. Klang association 22. Speech dissociation 23. Irrelevance 25. Blocking 26. Stereotype of speech 27. Echolalia	16.19 16.22 16.20, 16.21, 16.26, 16.27*, 16.28 1.14*, 16.30*, 17.09* 16.25 16.29 16.24
5. Affect-laden thoughts	28. Gloomy thoughts 29. Elated thoughts 30. Hopelessness 31. Suicidal thoughts	7.04, 7.05 9.1*, 9.2, 9.3, 9.4, 9.5, 9.7, 9.8 7.06, 7.07, 7.11 7.11*, 7.12
6. Predelusional signs	33. Delusional mood 34. Ideas of reference 35. Questions reason for being 37. Perplexity	11.04, 11.09, 11.10, 12.22, 15.27 12.17, 12.18 10.10 15.06*, 15.12
7. Experiences of control	38. Thought alienation 39. Thoughts spoken aloud 40. Delusions of control	12.01, 12.02 12.03, 12.04 12.09, 12.11, 12.12, 12.13
8. Delusions	41. Persecution 42. Guilt 43. Self-depreciation 44. Nihilistic 45. Grandeur 46. Reference 47. Presence of delusional system 48. Hypochondriacal 49. Special mission 50. Religious 51. Fantastic 52. Sexual 53. Impending doom	12.31*, 12.32*, 12.33, 12.34, 12.35, 12.36, 12.38, 12.39, 12.40* 12.46, 12.50, 12.51, 12.52 12.47, 12.48, 12.49 12.60, 12.61 12.65*, 12.67, 12.68, 12.69, 12.70, 12.71, 12.73 12.19, 12.20, 12.21, 12.23*, 12.24, 12.26 12.93* 12.58, 12.59 12.72 12.77 12.80 12.83 12.54

5 UAs were excluded because they did not fit well into any of the groups and it was considered unsuitable to create 5 new groups to accommodate them. These units were "perseveration", "frequent auditory hallucinations", "groaning", "loss of emotions", and "increased interest".

* Items in which only rating 2 was taken to signify presence.

TABLE 7.8 (continued) COMPOSITION OF GROUPS OF UNITS OF ANALYSIS

Groups	Units of Analysis	Items (indicated by PSE number)
9. Neurasthenic complaints	54. Obsessive thoughts 55. Worries 56. Lack of concentration 57. Memory difficulties 58. Hypochondriacal 59. Undecided 119. Decreased interest	10.01, 10.02, 10.03, 10.04, 10.05, 10.06, 10.07, 10.08, 10.09 1.05* 3.1*, 3.2*, 3.3* 3.4* 4.17 6.2* 2.1*
10. Lack of insight	60. Lack of insight	11.36*, 11.46*, 12.07*, 12.15*, 12.28*, 12.44*, 12.56*, 12.63*, 12.75*, 12.78*, 12.81*, 12.84*, 13.1
11. Distortion of self-perception	61. Changed appearance 63. Looking at self 64. Break of self-identity	11.12, 11.13, 11.17, 11.18 11.16 11.15
12. Derealization	62. Derealization 65. Distortion of time perception	11.01, 11.02, 11.03, 11.05, 11.07, 11.08 11.06
13. Auditory hallucinations	66. Presence of verbal hallucinations 67. Voices speak to patient 69. Nonverbal auditory hallucinations 70. Presence of auditory hallucinations	11.23 11.26, 11.32 11.34, 11.35 11.20
14. "Characteristic" hallucinations	68. Voices speak full sentences 72. Voices discussing patient 73. Hallucinations from body 74. Voices comment on patient's thoughts 75. Voices speak thoughts	11.27* 11.25 11.24 11.28, 11.31 11.30
15. Other hallucinations	76. Visual 77. Tactile 78. Olfactory 79. Sexual 80. Somatic 81. Gustatory	11.38 11.40 11.42 11.43 11.44 11.41
16. Pseudo-hallucinations	82. Auditory 83. Visual	11.21 11.39
17. Depressed-related	32. Special depression 84. Depressed mood 85. Observed elated mood	7.10 7.01*, 7.02*, 7.03*, 7.09*, 15.03*, 15.04*, 15.23* 15.05*, 15.22
18. Anxiety, tension, irritability	86. Morose mood 88. Irritability 89. Tension 90. Situation anxiety 91. Anxiety	15.24* 5.1*, 15.14*, 15.15 8.1*, 8.2*, 8.3* 8.7, 8.8 8.4*, 8.5*, 8.6*, 15.02*, 15.25*
19. Flatness	92. Flatness 93. Apathy	15.07*, 15.09 15.16

5 UAs were excluded because they did not fit well into any of the groups and it was considered unsuitable to create 5 new groups to accommodate them. These units were "perseveration", "frequent auditory hallucinations", "groaning", "loss of emotions", and "increased interest".

* Items in which only rating 2 was taken to signify presence.

TABLE 7.8 (continued) COMPOSITION OF GROUPS OF UNITS OF ANALYSIS

Groups	Units of Analysis	Items (indicated by PSE number)
20. Incongruity	95. Incongruity of affect	15.10
21. Other affective change	94. Ecstatic mood 97. Haughtiness 98. Ambivalence 101. Lability of affect 102. Ambitendence	15.26 15.17 15.20 15.21 14.35
22. Indication of personality change	8. Odd appearance and behaviour 103. Change of interest 104. Change of sex behaviour 105. Autism 106. Abnormal tidiness 110. Social withdrawal	14.04, 14.22 2.3 5.9 17.06 14.02 5.5*, 5.6*
23. Disregard for social norms	108. Disregard for norms 109. Self-neglect	14.23, 14.24*, 14.25, 14.26 14.03
24. Other behavioural	6. Talking to self 27. Disorder of pitch 96. Giggling to self 100. Demonstrative	14.31, 16.14, 16.15, 16.16, 16.17 16.05*, 16.06, 16.07, 16.08 14.20 15.11
25. Psycho-physical	111. Early waking 112. Worse in morning 113. Worse in evening 114. Diminished appetite 115. Sleep problems 116. Increased appetite 117. Increased libido 118. Decreased energy 120. Decreased libido 121. Constipation	4.12 4.14 4.15 4.01*, 4.03* 4.07*, 4.09*, 4.10*, 4.11* 4.02* 5.8 4.16* 5.7* 4.05*
26. Cooperation difficulties, circumstances related	125. Biological treatment 128. Environmental circumstances 129. Speech impediments	1.01, 1.02, 14.44 17.13, 17.14, 17.15, 17.16 16.03, 16.09
27. Cooperation difficulties, patient related	36. Suspiciousness 122. Suggestibility 123. Poor rapport 126. Unwilling to cooperate 127. Inadequate description	1.13*, 12.30, 15.18, 17.08 15.13 17.05 1.07, 1.16, 12.90, 17.04, 17.10 1.06*, 1.08, 1.09, 1.10, 1.11, 1.12, 1.13, 16.12, 17.07, 17.11

5 UAs were excluded because they did not fit well into any of the groups and it was considered unsuitable to create 5 new groups to accommodate them. These units were "perseveration", "frequent auditory hallucinations", "groaning", "loss of emotions", and "increased interest".

* Items in which only rating 2 was taken to signify presence.

APPLICABILITY AND RELIABILITY OF METHODS

Any method for evaluating patients cross-culturally must be applicable in different centres and must be capable of generating comparable information in the different settings in which it is used. For the IPSS, the semi-structured interview in which specific questions are suggested was selected for obtaining information on patients (see Chapter 5) because it combined a universally used evaluation technique — the interview — with a format standard enough to promote the collection of comparable data. The present chapter describes the applicability and reliability of this method and the schedules developed to implement it.

8.1 Applicability8.1.1 Applicability of the Present State Examination schedule

At the end of the data collection period, investigators were asked to comment on how applicable they felt the different schedules had been to the conditions and the patients in their centres. There was a unanimous acceptance of the Present State Examination (PSE) schedule as a valuable method for collecting present state information. In all FRCs psychiatrists found that this method approximated the techniques used in their centres prior to the IPSS for the evaluation of patients. For the most part the PSE was found to be easy to administer, the concepts and terminology were familiar, and the information sought was deemed generally relevant and appropriate.

It is especially important to stress the overall approval and acceptance of the PSE as method and schedule since it was used in nine vastly different psychiatric centres by psychiatrists with different backgrounds and orientations. It is also valuable to describe the anticipated and actual problems with the schedule.

It had been feared that such a procedure, especially considering the semi-structured nature of the interview, might exclude or make it extremely difficult to evaluate patients who were agitated, guarded, or withdrawn. In fact, this difficulty was not encountered. Very few patients in any of the centres were so agitated or otherwise difficult to interview that they could not be evaluated productively within the first few days of their admission. Since most of the FRCs were also the treatment centres for the patients, those few patients who were difficult to interview could generally be successfully reinterviewed within a few days. In the sample of 1202 patients seen, only twelve were reported to be totally unwilling to participate in the interview within the first week of admission. For these patients, the behaviour section of the PSE was completed and the schedule submitted to Headquarters for analysis with other data. The number of uninterviewable patients was smaller than expected. Consistent with this finding is Spitzer's using a similar type of interview, that only seven out of 600 patients seen

refused to complete the interview. However, it is also possible that there was an unsuspected subject selection factor operating in the FRCs although a broad spectrum of acutely psychotic patients was included in the study. Unwitting selection if it occurred was probably rare. For those patients whom it was somewhat difficult to interview because of moderate agitation, restriction of speech, or guardedness, and who gave incomplete information, the behaviour section of the PSE was especially helpful in describing the current psychiatric status.

Some difficulties were of course experienced in using the PSE as a common denominator for standardized evaluation. Several FRCs found the PSE longer than they would have liked, especially for some patients who became fatigued with the numerous questions, or for highly disturbed patients. This problem had been anticipated, but it had been decided not to shorten the form until actual experience in the IPSS could provide more guidance as to which of the items could be eliminated without compromising the goal of comprehensive evaluation of patients from a wide variety of cultures. Interestingly, many investigators found that the range of items in the schedule helped to increase the breadth of inquiry of the psychiatrists in the centres and to orient them towards aspects of symptomatology that they had previously tended to slight. The development of the schedule based on suggestions from several FRCs thus had the important result of increasing the scope of awareness of the psychiatrists involved in the project.

The frequency with which each of the different ratings was given within the entire patient sample was calculated in order to evaluate whether the PSE contained many items that were of questionable utility either because they were rated present so rarely or because they were rated present so frequently that they failed to distinguish between different types of patients. Table 8.1 lists the items rated 1, 2, or ? very rarely and the items rated 1, 2, or ? very frequently. It also indicates the frequency of ratings of 1, 2, or ? for a sample of other items of particular interest.

One specific limitation of the PSE was that centres such as Taipei and Washington with particular interest in patient psychodynamics found that the semi-structured schedule did not generate adequate information in this area. To offset this problem, it was suggested that in the future the mental status interview might be given in two or three parts so as to allow the investigators to detour occasionally from the interview and investigate further some of the interpersonal and intrapsychic aspects of the symptoms.

Another problem in the use of the PSE in the different FRCs — a problem that would occur with any cross-cultural method — is that certain concepts used in one culture were not readily translatable into others, (Chapter 6).

There was some concern that the use of semi-structured interviews might omit investigation of symptoms that were commonly found in some of the FRCs. Again, this fear did not materialize. Only a few symptom areas — delusions of jealousy and delusions of being possessed by a god or by witchcraft — needed to be expanded to make the schedule more suitable for those centres that frequently encountered patients with these symptoms.

Still another concern was whether the items of the PSE, which were supposed to represent psychiatric symptoms, would in fact reflect pathology in all of the cultures from which the samples were drawn. One centre, Ibadan

TABLE 8.1

FREQUENCY OF POSITIVE RATINGS
OF PSE ITEMS

<u>Items Rated 1,2, or ? with</u> <u>Frequency \leq 2%</u>	<u>Items Rated 1,2, or ? with</u> <u>Frequency $>$ 50%</u>
10.11 Other obsessional phenomena	1.04 Worrying which can't be stopped
11.19 Other deperson. or dereal.	1.05 Appropriateness of worries
14.10 Absence of voluntary movement	2.1 Loss of interests
14.31 Lips moving soundlessly	3.1 Diff. concentrating on actions
14.33 "Mitgehen"	4.07 Trouble sleeping
14.34 Imitation of examiner's movements	4.09 Difficulty getting off to sleep
14.41 Refusal to do what is asked of him	4.16 Less energy
16.03 Stammers or stutters	5.1 Gets angry with things or people
16.09 Malformations of speech apparatus	5.3 Shows anger
16.16 Talks to self or hall. incoherently	5.5 Wants to stay away from people
16.17 Talks to self or hall. coherently	7.01 Felt miserable or depressed
16.22 Rhymes	7.02 Cried
16.24 Echolalia	7.09 Variable mood
	12.89 Preoccupied with delusions
	12.92 Delusions widespread
	13.1 Overall rating of insight
	14.01 Behavioral abnormalities observed
<u>Items Rated 1,2, or ? with</u> <u>Frequency $>$ 2% $<$ 10%</u>	<u>Frequency of Rating 1,2, or ? for</u> <u>Other Items of Interest</u>
1.09 Fails to respond verbally	8.1 Anxiety 46.4%
1.16 Antagonistic, hostile, negativistic	11.20 Auditory hallucinations 36.8%
11.39 Visual hallucinations	15.07 Flattened affect 32.3%
11.45 Other hallucinations	
12.59 Feels body is decaying	
12.60 Part of body missing	
12.61 Feels he doesn't exist	
12.62 Other nihilistic delusions	
15.08 Other abnorm. of facial expression	
15.11 Exaggerated expression of feelings	
15.13 Suggestibility	
15.15 Outbursts of anger in interview	
15.17 Haughty, superior attitude	
15.19 Other abnormal reactions	
15.20 Ambivalence	
15.26 Exaltation, ecstatic mood	
15.28 Other abnormal affect	
17.02 Appeared confused in interview	
17.13 Membership in minority group	
17.14 Specific situation in hospital	
17.15 Difficult life situation	
17.16 Personality traits distorting report	
17.17 Other extenuating circumstances	

reported that some of the questions were either considered inconceivable, "eliciting bewilderment and laughter" (questions on homosexuality, depersonalization, and derealization), or reflected cultural beliefs rather than pathological features (some of the questions about "delusions"). Other FRCs encountered this kind of problem more rarely. In Washington three patients were seen who came from families or subcultures where hearing voices or feeling possessed by God was an acceptable "normal" experience that did not seem to represent a symptom for these patients. To some extent, however, the importance of this problem was reduced by the PSE instruction to rate phenomena only if they were considered pathological.

There had been concern early in the study over whether it would be possible for the investigators to modify their usual interview style enough to adhere to the PSE format, to follow the wording of the items fairly closely in asking the questions, and to make the ratings in the course of the interview. This latter requirement was essential since, with a minimum of 137 responses to specific questions to be rated, considerable accuracy would be lost if ratings were made after completing the interview. In general, the psychiatrists in all centres found it suitable to follow the format closely and also to rate while interviewing. The general acceptance of the semi-structured interview by the investigators is probably related to Spitzer's (1964) observation that once an interviewer has become familiar with an interview schedule he can follow it closely while still conducting the interview in a natural, flowing manner.

8.1.2 Applicability of psychiatric history and social description schedules

Obtaining psychiatric history and social data posed different problems from those encountered with the PSE. Early in the IPSS an attempt was made to construct history and social data forms that would, like the PSE, provide standardized semi-structured methods for eliciting and recording this data. However, such a procedure presented many formidable difficulties. One was that many different methods were commonly in use in the various centres for obtaining psychiatric history and social data. In a few centres psychiatrists obtained this data from patients. In others social workers, nurses, or students obtained data from interviews with relatives. In some FRCs information was obtained from notes from previous psychiatric contacts; in others no such notes were available. This variety in the usual methods and sources for obtaining psychiatric history information made it difficult to establish a standardized method for use in centres all having limited personnel. For this reason the idea of having fully standardized interviews for collecting history and social data had to be abandoned, and it was decided instead to use less structured forms for recording data.

Since the length of these forms sometimes made it difficult to complete them thoroughly, the early drafts of the Psychiatric History and Social Description schedule went through several revisions. The schedule had a broad focus in order to permit the identification of areas of information that would be most appropriate to obtain in the future. During Phase II, the investigators agreed that while the Psychiatric History and Social Description forms were useful and applicable, it would be important to develop more

standardized methods to obtain these kinds of data.

The Physical and Neurological Examination and Condition on Admission forms even more than the Psychiatric History and Social Description schedules, were in the form of worksheets; that is, they had items to be filled in from different sources of data rather than specifying exactly how the data were to be obtained. These were not difficult to complete, but the fact that a worksheet form had to be used reflected the continuing problems in developing more structured, standardized methods for collecting this type of information from the different Centres.

8.2 Reliability

The second question in assessing the methods of the IPSS for cross-cultural patient evaluation was whether the information recorded on the forms reflected patient characteristics in the same way in the different centres. Several studies (Beck et al. 1962, Palmer, 1943, Rosenzweig et al. 1961) have shown that variation in interviewing and rating techniques can cause considerable discrepancy in the collection of data. Although reliability is an important factor for any rating instrument, it was especially important in the IPSS since the cross-cultural difference between raters could have masked or exaggerated the differences in the meanings of the ratings. Lin (1969) and Sartorius et al. (1970) described the importance and ramifications of these problems and reported early data collected in the IPSS indicating that the methods were reliable.

The reliability of interview methods for collecting data can be evaluated either by assessing separately the elements of reliability — the interviewers, the schedules, the raters, and the respondents — or by testing these elements together to evaluate the overall reliability of the methods, e.g. by analysing the process of eliciting data and the process of rating responses. In view of pioneering goals and scope of the IPSS, an overall evaluation of reliability of the evaluation process in the different Centres was deemed most useful. This was accomplished by focusing on the stages of the data collection process and considering three main questions: first, were the forms utilized in similar ways in the different FRCs; second, did the raters agree in the way they recorded the responses, behaviour and other data about the patients being seen; and finally, because of the importance of stable measures at least over short periods of time for meaningful classification of patients, was there consistency of repeated evaluations so that two consecutive sets of ratings of a patient produced similar data. In the following section we shall take up each of these questions in turn. Again, because of its different form and content, the PSE will be discussed separately.

8.2.1 Reliability of the PSE

A great deal of comparability of use was built into the PSE since it is an interview schedule with precise instructions for interviewing the patients and was applied in each of the centres by psychiatrists trained in its use. In order to maximize further the similarity with which the interview was

administered, investigators interviewed patients during the exchanges of visits. This permitted observations and discussions of the interviewing techniques among the staffs of different FRCs, and led to increased agreement concerning exactly how the interview schedule was to be used. As a result of these exercises there was remarkably little variation among the investigators in the way in which they conducted the interview. Although stylistic differences were apparent among interviewers, the content of the questions asked and the inquiries made were remarkably similar. More quantitative measures of the similarity of interview administration were not attempted at this time.

The second problem of reliability was to determine to what extent raters scored similarly the responses and behaviours of the patients being seen. Wing et al. (1967) and Kendell et al. (1968) reported high reliability for earlier versions of the PSE when used by a smaller group of investigators from one FRC. In a series of reliability evaluations they found that while the general level of reliability was high, ratings based on patient reports were more reliable than ratings based on observations of behaviour. They also found that reliability of consecutive interviews was somewhat lower than that based on ratings made by two raters scoring one interview.

In the IPSS three methods for evaluating PSE reliability were used: (1) simultaneous ratings made by two or more investigators attending and rating an interview; (2) ratings made in consecutive interviews of the same patient; and (3) the rating of English language videotaped or filmed interviews by a large number of investigators from all Centres. The first two techniques were used for measuring "intracentre reliability", i.e. for comparing ratings between investigators working in the same FRC. The third technique, although technically more complicated and therefore less frequently used, was valuable for measurement of "intercentre reliability", i.e. for comparing ratings among investigators from different FRCs.

In testing the reliability of such evaluation procedures, it is important to decide what levels of reliability will be acceptable. A level of significance (for example $p = .05$) is often taken as an acceptable level for a statistical result; but as Spitzer and Cohen (1968) have pointed out, this significance test is inadequate. In studies with large numbers of subjects such as the IPSS, p can be very significant when, in fact, the actual level of agreement among raters is low. This is possible because with a large number of comparisons of raters a product-moment correlation coefficient (r) of .30 can be significant in terms of the small likelihood of deviations from 0 while the r value itself indicates that only 9% of the variance of the ratings made is shared by the raters. Even with $r = .70$, only 49% of the variance of ratings will be a function of the qualities of the patient, irrespective of the particular rater. For this reason and because a goal of the IPSS was to develop instruments ensuring that investigators in different centres would be describing patients in essentially the same way, it was important to see a high level of reliability, as well as of significance. Although no single figure for r can be supported definitely, a reliability level for these schedules that accounts for over 50% of the variance is not unreasonable. Level of agreement lower than this are important but they do not strongly support the utility of the schedules for making comparable descrip-

tions of patients.

Choice of statistical methods for data analysis. Before the data obtained from these interviews could be evaluated, it was necessary to select methods of data analysis based on the nature of the PSE ratings. Since the PSE attempts to adhere closely to the kinds of patient evaluation made clinically, many choices have to be made in transferring the ratings to data categories suitable for analysis. Must one require that two raters rate identically in order for the particular item to be counted as an agreement, or can the rating of "1" by one and "2" by another rater be counted as an agreement since both "1" and "2" ratings indicate that the symptom is considered present? If ratings are combined, is a "?" equivalent to a "0" or to a positive rating, or does "2" require a separate category? In the data analyses performed, several of these alternatives were employed in order to utilize a variety of approaches.

Another decision to be made was how best to compensate for the lack of variation of ratings for certain rarely occurring symptoms and for the non-normal distribution of ratings, since the overwhelming majority of ratings for any patient are "0" entailing a high probability of 0-0 agreements between two raters. Although some techniques of data analysis seemed to solve these problems better than others, more than one method was used for many of the analyses to permit a comparison of the results. The various tests of reliability and bias performed are summarized in Table 8.2.

In choosing between the various statistical techniques for analysing reliability, it was decided that the intraclass correlation coefficient (Hays, (1967)--- a one-way analysis of variance --- would involve the fewest assumptions while being optimally suited to the nature of the data. The formula for this statistic is

$$\underline{R} = \frac{S_b^2 - S_w^2}{S_b^2 + (m-1) S_w^2}$$

where S_b^2 is the variance between groups, S_w^2 is the variance within groups, and m is the number of ratings made on a particular patient for the particular unit of analysis or item under consideration. As an analysis of variance, the intraclass correlation coefficient measures which part of the total variance of a score is attributable to one particular source of variance. Used as a test of reliability where several raters rate many subjects, it measures that part of the variance of ratings on an item or group of items that is due to disagreements between raters, as compared with the total variance due to differences between patients as well as between raters and other sources. Thus, if paired raters score a series of subjects, intraclass \underline{R} considers the difference within each pair of ratings and compares them with the total variance for all ratings on all patients by the formula given. The maximum possible value for intraclass \underline{R} is 1, in those cases where there

TABLE 8.2 ANALYSES PERFORMED FOR STUDY OF
RELIABILITY AND BIAS OF PSE DATA

Source of Data	Form of Data Analysed	Statistic Used	Main Finding
Intracentre simultaneous interviews	Items	Intraclass \underline{R}	Range of \underline{R} = .43 - .97 Median \underline{R} = .77
"	Units of Analysis (UAs)	"	Range of \underline{R} = .47-.96 Median \underline{R} = .81
"	UAs (dichotomized)	"	Median \underline{R} = .74
"	"	% of Total Agreement	> 90% for most UAs
"	"	% of Serious Disagreement	Range = 7 - 53%
"	"	% Agreement on Positive Ratings	Generally High, but 21 UAs < 50%
"	27 Groups of Units	Intraclass \underline{R}	Mean \underline{R} = .83
First Five and last Five Simultaneous Interviews	"	"	Little Change
Intercentre Reliability Interviews	UAs	Intraclass \underline{R}	Range of \underline{R} = .00 - .84 Median \underline{R} = .45
Intercentre Reliability	27 Groups of Units	"	Range of \underline{R} = .00 - .87 Median \underline{R} = .57
Reliability of Raters from same Centre in inter-centre exercises	"	"	Range of \underline{R} = .23 - .90 Median \underline{R} = .66
"	"	Bias Ratio (Rater and Centre)	Minimal Bias
"	"	Bias Ratio (Raters developed vs Raters developing countries)	Minimal Bias
"	"	Bias Ratio (UK trained compared with US trained)	No significant difference
"	5 Groups of Symptoms	Index of Bias	Evidence of Rater and Centre Bias
Intracentre Consecutive Interviews	27 Groups of Units	Intraclass \underline{R}	Range of \underline{R} = .10 - .85 Mean \underline{R} = .57

is complete agreement among raters. The minimum possible value in cases of complete disagreement is 0. A positive value for \underline{R} is roughly interpretable as the percentage of variance of a measure attributable to changes in the object being measured. One characteristic of this statistic is that if there is little or no variance in the measure, e.g., if all ratings for an item are 0 the statistic is meaningless.

The intraclass correlation coefficient yields two results, an "average" \underline{R} and a "single" \underline{R} . The former is used primarily to compare an average of a group of ratings with the average of another group of ratings made on the same patient; the latter gives a measure of the likelihood that a second rating on a patient will be similar to the first. The single \underline{R} is lower in value and gives a more conservative estimate of reliability, but was more appropriate for the purpose of the IPSS. For the analyses to be described here, the "single" \underline{R} was used.

The intraclass correlation coefficient has several useful characteristics. It does not require dichotomous ratings and thus does not require the collapsing of categories. Unlike the product-moment \underline{r} , it can be used with more than two raters; also unlike the product-moment \underline{r} where $\underline{r}=1$, if one rater rates at a constantly higher level than the others, it gives a measure specifically of agreement rather than merely a measure of any constant relationship. Weighted kappa (Cohen, 1968) a statistic used to measure the reliability of nominal data, was not used because although compared to \underline{r} it has the advantage of compensating optimally for chance agreement in skewed data, it does not have the more important advantage possessed by intraclass \underline{R} of utilizing the ordinal qualities of the PSE data.

To determine intraclass \underline{R} for reliability of items, ratings were assigned the following values: 0=0, ?= $\frac{1}{2}$, 1=1, 2=2. Ratings of NR, NA, and NI were excluded (see Chapter 5). The decision not to assign a value to ratings of NR, NA, and NI in these analyses was made because they are of a different nature from 0, ?, 1 and 2, and because none of them gives an independent measure of whether the patient has or does not have a particular symptom. It is the description of what symptoms a patient has that was most crucial for the goals of the IPSS.

Results of reliability tests. To test the intracentre reliability of the PSE, an average of 21 interviews rated simultaneously by two psychiatrists was conducted in each FRC. A total of 190 of these interviews were performed.

(1) Intracentre reliability of items

The intracentre reliability of the individual PSE items was analysed for all interviews with simultaneous ratings from all of the Centres. The range of values of intraclass \underline{R} for the 360 items was from $\underline{R}=.97$ to $\underline{R}=.43$ with a median of $\underline{R}=.77$. In general these values are evidence for the acceptable reliability of the items when rated by investigators from the same Centres.

A striking finding was the high proportion of items rated from patient reports (49 out of 50) among the items with the highest reliability, and the

high proportion of ratings based on observation of patient behaviour rather than on patient reports (59 out of 61) among the items with the lowest reliability. Rather than describe in more detail the results of item reliability, it was decided to investigate the reliability of the combinations of items as they occur in Units of Analysis, since these units were more meaningful both statistically and clinically.

(2) Intracentre reliability of Units of Analysis

To evaluate the reliability of the 124 Units of Analysis (UAs), which were more practical to use than the 360 items for analysing profiles and patient characteristics, an intraclass correlation was performed on the simultaneous ratings made in each centre for each UA. The formula used for score of a UA was the sum of the value of the ratings received on the items in the unit divided by the maximum possible score for the UA (two times the number of ratings made, excluding any items rated NI, NA, or NR). This formula limited the range of scores possible for any UA to .0 - 1.0. The pooled intraclass \bar{R} , the method for obtaining a kind of average \bar{R} (actually a log-weighted average \bar{R}), was calculated for each UA by using data from all centres in which there was more than the minimal variance of one positive rating for that UA. The results for the ten UAs with the highest and the ten UAs with the lowest reliability and the median UA are reported in Table 8.3.

The results demonstrate that many UAs have extremely high reliability by pooled \bar{R} with a very acceptable median pooled $\bar{R}=.81$. The data also demonstrate that the range of reliability of the UAs is considerable when the values from individual centres are considered separately.

There were nine UAs with very little or no variation in scores that were never or only once rated positively. These were units dealing mainly with observed behaviour and speech (stupor, waxy flexibility, echolalia, compliance, stereotypy, ambitendence, and abnormal tidiness) with only two exceptions (the UAs for delusions of special mission and for looking at self). These invariant UAs should be considered separately from the rest of the data in evaluating IPSS results since their reliability is indeterminate.

Other methods of analysing the intracentre reliability of the UAs were used to see whether they would give similar results. In general they supported the overall acceptable reliability of the schedules but also indicated how dependent the more detailed findings are on the specific statistical methods used. In one such method, the intracentre reliability of UAs was analysed again using intraclass \bar{R} but dichotomizing all ratings into presence-absence categories (NA, 0, ? = absent; 1, 2 = present) and considering a UA as "present" if any item in it were rated present. The use of dichotomized as compared with the scaled ratings resulted in a lowering of the reliability of many UAs and a decrease in the median \bar{R} from .81 to .74. Using the dichotomized rather than scaled measures also caused marked rearrangement of the rank order of reliability of the UAs so that most units changed by at least 10 ranks, many by 40-60 ranks and a few by over 100 ranks out of 124.

Reliability was also assessed using other measures, on a somewhat modified earlier version of the UAs. Ratings were dichotomized so that ratings of 0, ?, NA, NR, and NI were considered to indicate absence of symptoms and ratings

TABLE 8.3

INTRACENTRE RELIABILITY OF 10 UNITS OF ANALYSIS
WITH THE HIGHEST R AND 10 UNITS OF ANALYSIS WITH LOWEST R

Units of Analysis	Pooled <u>R</u> Across 9 Centres	Median <u>R</u> Across 9 Centres	Range of <u>R</u> Across 9 Centres		
			Highest	Lowest	
	Suicidal	0.96	0.92	1.00	0.17
H	Elated Thoughts	0.95	0.96	0.99	0.54
I	Ideas of Reference	0.95	0.93	0.99	0.72
G	Delusions of Grandeur	0.95	0.89	0.99	0.48
H	Thoughts Aloud	0.95	0.93	1.00	0.58
E	Derealization	0.94	0.94	1.00	0.67
S	Lack of Concentration	0.94	0.94	0.98	0.62
T	Hopelessness	0.94	0.92	0.98	0.41
	Delusions of Persecution	0.94	0.90	1.00	0.79
	Delusions of Reference	0.93	0.94	0.96	0.78
	Median Unit: Break of Self-identity	0.81	0.78	0.87	0.69
	Change of Interest	0.62	0.42	0.90	0.09
	Speech Dissociation	0.61	0.46	0.85	0.14
L	Perplexity	0.61	0.58	0.81	0.17
O	Lability	0.60	0.47	0.87	0.24
W	Stereotypes of Behaviour	0.59	0.56	0.66	0.47
E	Increased Libido	0.59	0.53	0.73	0.26
S	Negativism	0.57	0.54	0.64	0.44
T	Perseveration	0.54	0.55	0.63	0.35
	Hallucinations from Body	0.53	0.46	0.64	0.27
	Morose Mood	0.47	0.53	0.63	0.21

of 1 and 2 were considered to indicate presence of symptoms (Sartorius, et al. 1970; WHO, 1969). The first measure, Index of Total Agreement (all agreements on presence of a symptom plus all agreements on absence of a symptom divided by the number of pairs of ratings of the symptom), was calculated for all UAs and was generally very high (90% or more). This high score was partly determined by the high proportion of ratings of absence of symptoms, which increased the probability of agreement. For this reason other measures of agreement were also used. The Index of Serious Disagreement (the number of disagreements divided by the number of pairs with at least one positive rating), which gives a measure of the percentage of disagreements per positive rating was calculated for each UA and showed a range of values from 7% to 53% for different UAs. This range is similar to that found in the joint United States-United Kingdom Diagnostic Project (Kendell et al., 1968) using generally similar interview schedules and the same statistical procedures. A third measure was used, again to overcome possible undue weighting from the large number of symptoms rated absent. This measure was the Percentage Agreement on Presence of Symptoms: the number of agreements on presence of a symptom divided by the number of pairs of ratings of the symptom with at least one rating of presence. Calculated for all UAs it demonstrated fairly good agreement for the majority of units. In 48 UAs there was over 70% agreement; in 55 UAs there was a range from 50-70%; and in 21 UAs the percentage of positive agreement was less than 50%. Many of these lowest UAs had been rated positively only rarely.

When calculated on a patient-by-patient basis rather than on a unit-by-unit basis for a sample of 110 patients seen early in the study from all centres on whom simultaneous ratings had been made, the percentage agreement on presence of symptoms varied from 9% on some patients to 100% on others. The average agreement on presence for all the UAs for a given centre varied from 49% to 91%. Both the lowest and highest agreements tended to be for those patients with few positive ratings. This interrelation between the statistic used and the nature of the data indicated that this measure of agreement when applied on a patient-by-patient basis was less meaningful for patients with few symptoms. Nevertheless with this method too, the mean percentage agreement on positive ratings was 66%.

To evaluate the reasons for consistently lower reliability of certain UAs, the data were analysed to see whether decreased reliability was predominantly a function of such mechanical factors as the number of items in a UA or the number of positive ratings assigned to a UA. There was a suggestion of a relationship between the number of items in a UA and the reliability of the unit but the structure of the data (124 UAs with a maximum of 13 items per unit) did not permit a statistical analysis. There was a small but significant positive correlation (product-moment $r = .36$, $p < .005$) between the reliability of a UA and the number of pairs of ratings with at least one positive score for that unit. This indicates that a significant but small amount (13%) of the variance of reliability of a UA was accounted for by the amount of positive use.

Another characteristic that differentiated the more reliable from the less reliable UAs was that the more reliable units generally consisted of items describing symptoms rated on the basis of patients' reports while

those with lower reliability consisted of behaviour items rated from observation. Kendell et al., (1968) and Lin (1969), also found ratings of observed behaviour to be less reliable than ratings of symptoms reported by patients. This may be related to the fact that behaviour ratings are made on the basis of observation during an entire interview, whereas rating an answer to a question is a more circumscribed event. It may also be that ratings of observed behaviour are more often dependent on the norms of behaviour held by a given psychiatrist than are the topics covered by specific questions asked of the patient.

The different methods of analysing the data on intracentre reliability all indicate that the general intracentre reliability of the PSE is high. These results concur with similar findings of Wing et al., (1967) and Kendell et al., (1968), who worked with an earlier version of this schedule. More specific conclusions were less certain. The rank order of reliability of UAs varied greatly depending on the statistical technique used, and the range of reliability for individual UAs varied considerably among different Centres. The intercentre reliability of UAs will be discussed below.

(3) Intracentre reliability of Groups of Units of Analysis

A final test of intracentre reliability of the PSE was performed using 27 Groups of Units of Analysis (GUAs) in order to condense the 124 UAs into a smaller more manageable number of groups, each of which would be comprised of more items than the UAs.

A full list of Groups of Units of Analysis and the units and items that compose them is given in Chapter 7. The formula used for the score of a group was the sum of the value of the ratings received on the items in the group divided by the maximum possible score for the group. The mean intraclass R of these groups was high (.83).

The intracentre reliability for GUAs as for UAs revealed many differences in the reliability of ratings among the different Centres. The mean reliability for all GUAs together for each FRC varied from intraclass $R=.53$ to $R=.90$, with individual GUAs varying in reliability from Centre to Centre as much as from $R=.00$ to $R=1.00$ for the same group (Table 8.4). This amount of variation in reliability again serves as a warning against too detailed an interpretation of these data, especially when applied to subsamples or to data that might have only slight variance. It is also significant that those Groups of Units comprised mainly of items based on patient reports rather than rater observation had the highest reliability and a narrower range of variation from Centre to Centre. Again, this parallels the results of the analysis of reliability of the UAs.

One question raised was to what extent the intracentre reliability of the PSE was a consistent function of the particular FRC in which the data was collected. Table 8.5 presents reliability data by centre and shows that the mean reliability differed from one FRC to the next, with the highest mean R for all GUAs in the Field Research Centres in Moscow (.90) and Agra (.86) and the lowest in Cali (.53), and Washington (.54). These differences in reliability for each FRC were generally noted regardless of whether the data were analysed in terms of items, UAs or GUAs and regard-

TABLE 8.4

27 GROUPS OF UNITS OF ANALYSIS
RANKED IN ORDER OF
INTRACENTRE RELIABILITY

Groups of Units of Analysis	Pooled R Over 9 Centres	Range over 9 Centres		R/O*	No. of Items
		High	Low		
Auditory Hallucinations	0.96	0.99	0.48	R	5
Experience of Control	0.96	1.00	0.73	R	8
Derealization	0.95	1.00	0.63	R	7
Delusions	0.95	0.98	0.83	R	39
Other Hallucinations	0.95	0.98	0.56	R	6
Neurasthenic Complaints	0.92	0.98	0.68	R	17
Affect Laden Thoughts	0.91	0.97	0.65	R	14
Psychophysiological	0.90	0.97	0.34	R	14
Distortion of Self-perception	0.89	0.98	0.36	R	6
Anxiety-Tension-Irritability	0.89	0.96	0.65	R	14
"Characteristic" Hallucinations	0.87	0.95	0.37	R	6
Depressed-Elated	0.87	0.95	0.39	1/2	10
Qualitative Psychomotor Disorder	0.86	1.00	0.00	O	10
Predelusional	0.84	0.95	0.66	R	10
Other Behaviour Change	0.84	0.96	0.46	O	11
Odd Behaviour and Withdrawal	0.83	0.93	0.51	1/2	8
Quantitative Psychomotor Disorder	0.82	0.99	0.00	O	16
Cooperation Difficulties, Patient related	0.81	0.96	0.10	O	8
Qualitative Disorder of Form of Thinking	0.79	0.99	0.02	O	13
Lack of Insight	0.77	0.86	0.53	R	13
Disregard for Social Norms	0.77	0.87	0.00	O	5
Pseudohallucinations	0.76	1.00	0.00	R	2
Quantitative Disorder of Form of Thinking	0.73	0.83	0.00	O	5
Other Affective Change	0.72	0.81	0.00	O	5
Flatness	0.72	0.89	0.06	O	3
Incongruity	0.65	0.82	0.00	O	1
Cooperation Difficulties, Circumstances Related	0.60	0.77	0.00	O	3

- * R = Contains mostly reported items
 O = Contains mostly observed items
 1/2 = Contains both reported and observed items

TABLE 8.5

INTRACENTRE RELIABILITY OF GROUPS OF UNITS OF ANALYSIS: INTRACLASST CORRELATION COEFFICIENT (R)

GROUPS	AARH	AGRA	CALI	IBAD	LOND	MOSC	TAIP	WASH	PRAG
1. Qualitative Disorder of Form of Thinking	0.02	0.80	0.73	0.19	0.47	0.99	0.49	0.59	0.83
2. Experience of Control	0.91	0.99	0.84	0.97	0.95	0.85	1.00	0.73	0.83
3. Delusions	0.94	0.98	0.83	0.88	0.98	0.92	0.98	0.92	0.93
4. Lack of Insight	0.81	0.73	0.53	0.72	0.71	0.75	0.84	0.62	0.86
5. Auditory Hallucinations	0.88	0.98	0.99	0.97	0.95	0.97	0.98	0.48	0.96
6. Flatness	0.12	0.89	0.60	0.06	0.66	0.89	0.83	0.10	0.77
7. Incongruity of Affect	0.72	0.73	0.44	0.25	0.00	0.82	0.69	0.10	0.69
8. Odd Behaviour and Withdrawal	0.90	0.77	0.62	0.52	0.51	0.89	0.93	0.76	0.89
9. Other 19 Groups Combined	0.66	0.85	0.47	0.58	0.70	0.90	0.80	0.54	0.80
Mean for 27 Groups	0.64	0.86	0.53	0.58	0.69	0.90	0.82	0.54	0.81

less of whether the data were scaled or dichotomized. There were many possible sources for these differences in rater reliability, including the previous training of the raters, the amount and type of experience the pair of psychiatrists had in working together and the amount of time they spent comparing ratings after conducting jointly rated interviews.

Other sources of the differences in levels of reliability of the GUAs were also investigated. The reliability of the individual groups was compared with the number of items per group. The product-moment r was significant ($r=.47$, $p < .005$), supporting Lorr's contention (1971) that ratings of single items representing psychiatric symptoms are often unreliable and that it is therefore often helpful to use many items related to one area so as to improve reliability. In the IPSS the lower reliability of groups concerned with observed behaviour as compared with groups concerned with patients' reports of symptoms may thus be partly due to the fact that there were fewer items in such groups and that such items were rarely rated positive. Future work to improve the reliability of these groups will have to take these possibilities into consideration.

One of the variables (more a function of the centres and raters than the items) considered as a possible source of difference in reliability in the PSE was the amount of experience raters had in using the schedule. A rough test of experience was made by determining whether raters tended to agree more about their ratings later in the study, after a period of using the schedules and comparing notes with one another, than they did at the beginning of the study. This comparison was performed with the first five and last five simultaneous interviews from each Centre for the 27 GUAs. The results showed that reliability did increase for 20 of the 27 groups, but in only four areas was the change statistically significant ($p < .05$). A more controlled study would be necessary to determine precisely what kind and length of experience with the schedules could increase the reliability further.

(4) Inter-centre reliability of Units of Analysis

The next task in the study of PSE reliability in the IPSS was to evaluate the level of agreement on PSE ratings between psychiatrists from different Centres. This was of major importance since high intercentre reliability was crucial for the comparability of the IPSS methods. For this purpose 21 interviews held in different FRCs were rated live or from videotapes or films by psychiatrists from the different Centres. On average, each of these 21 interviews was rated by 10 psychiatrists resulting in 202 sets of ratings. The reliability exercises are continuing in the follow-up phase and will be reported in Volume II of the report of the IPSS.

The intercentre reliability of these interviews for individual UAs ranged from .00 to .84, with median $R=.45$. Five of the 124 UAs were never or only once rated positive (waxy flexibility, mutism, echolalia, questions reasons for being, and groaning) and their reliability was indeterminate. Those UAs with $R=.00$ were rated positively only rarely. Once again, the UAs with lowest reliability were those rated on the basis of observed behaviour. There was a generally low level of reliability for many UAs having just a few items, again recalling Lorr's warning about the low reliability of individual ratings.

(5) Intercentre reliability of Groups of Units of Analysis

To determine whether the lower level of reliability for many UAs was primarily a function of their having few items, the intercentre reliability of GUAs was calculated. The intercentre reliability of the groups is shown in the first data column of Table 8.6 along with the intracentre \bar{R} and consecutive interview \bar{R} , to be discussed later, for comparison.

In fact, intercentre reliability was appreciably lower than intracentre reliability for each group as well as for the mean and median of all groups together. The intercentre reliability data show that no group has $\bar{R} > .90$, only two groups have $\bar{R} > .80$, and a total of 4 have $\bar{R} > .70$. Seven have $\bar{R} \leq .40$. Once again, the groups containing items rated from patients' reports of symptoms have generally higher reliability than those based on ratings of observed behaviour (product-moment $r = .53$, $p < .01$.) Besides corroborating the findings discussed above, the higher reliability of the group of reported symptoms renders unlikely the possible explanation that low intercentre reliability was a function of language difficulties created by asking psychiatrists to rate interviews not conducted in their mother tongue. It is noteworthy that the groups important to the diagnosis of schizophrenia spanned all levels of reliability, with hallucinations and delusions being among the more reliable and key behavioural signs among the least reliable.

In order to determine whether the differences between the intracentre and intercentre reliability data were related to the different interview observation methods used, the ratings of filmed interviews used to evaluate intercentre reliability were also analysed to measure intracentre reliability. This was done by calculating intraclass \bar{R} separately for the pairs and triads of raters from the same FRC who rated the filmed interviews and then pooling these intracentre reliability values. The results were very similar to those for the intercentre values, with the range of \bar{R} for the GUAs varying from a high of $\bar{R} = .90$ to a low of $\bar{R} = .23$, with a median of $\bar{R} = .66$. Ratings based on patients' reports were again more reliable than ratings based on observations of patient behaviour. This strongly suggests that much of the lower intercentre reliability stemmed from the interview observation situation, possibly being due to difficulties in appreciating the context of the interview or in maintaining attention while rating filmed interviews, but that it was not primarily caused by language problems or specifically by intercentre disagreement. The important differences may have involved both increased difficulty in following and rating filmed interviews, and the lack of possibly unrecognized helpful rating cues available only in the rater-observer situation used to evaluate intracentre reliability.

Other potential sources of the lower reliability of the intercentre ratings were evaluated as well. To see whether the group of 7 raters trained in the United Kingdom rated more similarly than the entire group of raters, and whether the same might be true for the 6 raters trained in the United States, the reliability of each of these groups was calculated separately for the 27 GUAs. Although the reliability of these groups' ratings of the intercentre exercises were slightly higher than the reliability for the entire group of raters from all FRCs, a t test indicated that the difference was not statistically significant. A similar analysis dividing the raters into 3 groups

TABLE 8.6

RELIABILITY OF RATINGS FROM SIMULTANEOUS AND CONSECUTIVE
INTERVIEWS RANKED IN ORDER OF INTERCENTRE R

Groups of Units of Analysis	R/O*	Multiple Intercentre <u>R</u>	Simultaneous Intracentre <u>R</u>	Consecutive Intracentre <u>R</u>
Auditory Hallucinations	R	0.87	0.96	0.73
Other Hallucinations	R	0.80	0.95	0.85
"Characteristic" Hallucinations	R	0.78	0.87	0.41
Neurasthenic Complaints	R	0.73	0.92	0.33
Experience of Control	R	0.67	0.96	0.41
Psychophysiological	R	0.66	0.90	0.66
Qualitative Psychomotor Disorder	O	0.66	0.86	0.68
Derealization	R	0.63	0.95	0.70
Predelusional	R	0.63	0.84	0.62
Delusions	R	0.61	0.95	0.79
Qualitative Disorder of Form of Thinking	O	0.61	0.79	0.68
Depressed-Elated	1/2	0.60	0.87	0.60
Quantitative Disorder of Form of Thinking	O	0.59	0.73	0.66
Affect Laden Thoughts	R	0.57	0.91	0.73
Anxiety-Tension-Irritability	R	0.54	0.88	0.61
Cooperation Difficulties, Patient Related	O	0.48	0.81	0.57
Distortion of Self-perception	R	0.46	0.89	0.10
Odd Behaviour and Withdrawal	1/2	0.44	0.83	0.66
Other Affective Change	O	0.41	0.72	0.66
Pseudohallucinations	R	0.32	0.75	0.62
Quantitative Psychomotor Disorder	O	0.27	0.82	0.59
Flatness	O	0.27	0.72	0.79
Other Behaviour Change	O	0.25	0.84	0.30
Incongruity	O	0.16	0.65	0.74
Lack of Insight	R	0.08	0.77	0.60
Disregard for Social Norms	O	0.00	0.77	0.45
Cooperation Difficulties, Circumstances Related	O	0.00	0.60	0.29
Median <u>R</u>		0.57	0.84	0.62
Mean <u>R</u>		0.49	0.83	0.57

* R = Contains mostly reported items; O = Contains mostly observed items; 1/2 = Contains both reported and observed items

based on their amount of experience with the English language also failed to disclose significant differences in reliability among the groups.

Another possibility was that raters from countries with similar levels of development might rate more similarly because of possibly similar experiences. To evaluate this possibility the intercentre reliability data were calculated first by comparing all the ratings made by investigators from developing countries and then by comparing all the ratings made by investigators from developed countries. If different experiences contributed to lower reliability of ratings, one would expect that separating the ratings into these two groups would increase the reliability of each above that achieved by all raters grouped together. In fact this was only true for four of the 27 groups of units and in these the increase in reliability gained by dividing the raters in this way was minimal.

(6) Rater and center bias

A possible source of the lower intercentre reliability might be consistent rater or centre bias in making ratings. Katz and his co-workers (1969a) studying sources of unreliability and bias in psychiatric evaluation concluded that the diagnoses made by psychiatrists are closely related to the manner in which they perceive and rate symptoms. They demonstrated that psychiatrists from different countries rating the same filmed interview rated different levels of patient symptomatology. Sandifer et al. (1969) also found significant bias in psychiatrists from different countries rating the same interview.

An evaluation of possible rater and centre bias was carried out for the IPSS data to see if similar effects were present in this study. Again, the results of such an inquiry depend in part on the measures used. After trying several different techniques, it was decided that a ratio was the most useful statistic. A "bias ratio" was calculated, defined as

$$\frac{\bar{X}_{ij}}{\bar{X}_j}$$

i.e., the average rating (\bar{X}) for particular items made by rater i for patient j divided by the mean of the ratings of all raters of these items for the same patient. In this way a score greater than 1.0 for a particular rater or subgroup of raters would indicate a tendency on their part to rate that area of pathology higher than the average. A score of less than 1.0 would indicate that the rater or subgroup of raters under consideration tended to rate less pathology. From this ratio a weighted mean bias (B) for rater i was then calculated, where

$$B = \frac{aB_1 + bB_2 + \dots + B_n}{a+b + \dots + x}$$

in which a_i , b_i , x_i are the number of comparison raters rating each of the interviews ($1, 2, \dots, n$) rated by rater i . The weighted mean bias can be interpreted as a proportion. For example, $B=2$ indicates that a rater tends to rate twice as high as the average rater. $B=0.5$ indicates that a rater tends to rate half as high as the average rater. The same interviews used for the study of intercentre reliability were used for this analysis. The bias ratio of particular GUAs for raters who had rated those UAs in only one multiple interview were highly skewed so that their ratings for those groups were excluded from the analysis.

For individual raters, no rater showed a consistent trend to rate either more or less pathology than average for all or most groups. In fact, the mean bias ratio per rater for all groups together ranged only from a low of 0.7 to a high of 1.6, with a standard deviation of .23. This contrasted with a range of as much as 0.0-7.9 with a standard deviation of .87 for the bias ratios of raters on some of the areas taken individually. In fact, 13 out of the 24 raters had a mean bias ratio between .9 and 1.1, and 17 raters out of the 24 had a mean bias ratio between .8 and 1.2.

When ratings of the investigators from each FRC were grouped together and analysed as centre bias, the bias of individual raters was mitigated so that bias by centres was considerably less than bias by individual raters. Table 8.7 shows bias by centre for selected GUAs. Some groups (delusions, and hallucinations) seem less prone to bias than others. The range of mean bias by centre was 0.8 to 1.6. The higher value represented the ratings of the centre for which the ratings of only one investigator were included. The highest mean bias value for the rest of the centres was 1.2.

To test whether rating consistently more or less pathology was accounted for by a rater's coming from a particular FRC, a one-way analysis of variance was performed on these data. The results were not significant, indicating that this was not a significant factor in rater bias. In another analysis of bias ratio, all the bias ratios of investigators from the developing centres were compared with the bias ratios of investigators from the developed centres to see if this way of dividing the data would demonstrate more consistent patterns of bias. To accomplish this the arithmetical mean of the bias ratio from the four developing countries was compared to the mean from the five developed countries. The difference between the means (t test) was not significant. There were only minor differences as well for the individual groups of units.

A further analysis of bias was performed to see if raters trained in the same psychiatric centre tended to have similar rating bias. For this purpose the six raters who received a significant part of their psychiatric training in the United Kingdom were considered as one group and six raters who received a considerable part of their psychiatric training in the United States were considered as another group. A t test for significance of the difference of the mean bias between these groups demonstrated that the difference was not significant. Similarly, there was no sizeable difference between the two groups of raters on the individual GUA.

In a different approach to evaluating rater and centre bias, an analysis was performed using all simultaneous, multiple, and videotaped ratings from both Phase I and Phase II (Sartorius et al., 1970). For this purpose the

TABLE 8.7

BIAS* BY CENTRE IN RATING
GROUPS OF UNITS OF ANALYSIS

GROUPS	CENTRE								
	AARH	AGRA	CALI	IBAD	LOND	MOSC**	TAIP	WASH	PRAG
1. Quantitative & Qualitative Disorder of Form of Thinking (2 Groups Combined)	0.7	0.9	<u>0.6</u>	<u>0.6</u>	<u>0.6</u>	<u>1.8</u>	<u>1.6</u>	1.1	1.3
2. Delusions	1.0	1.2	0.9	1.0	1.2	1.1	1.0	1.2	<u>0.5</u>
3. Hallucinations & Pseudohallucinations (4 Groups Combined)	<u>1.4</u>	0.9	0.8	0.7	0.8	<u>2.8</u>	0.8	0.9	0.9
4. Depressed-Elated	1.3	0.8	<u>0.6</u>	<u>0.6</u>	<u>0.4</u>	1.2	0.8	<u>1.5</u>	0.9
5. Anxiety-Tension-Irritability	<u>2.7</u>	1.3	0.7	<u>0.2</u>	<u>0.6</u>	<u>0.2</u>	1.0	1.2	1.2
6. Flatness	0.9	1.0	<u>0.6</u>	0.9	<u>2.2</u>	0.9	<u>1.4</u>	0.7	<u>0.6</u>
7. Disregard for Social Norms	<u>0.4</u>	<u>1.5</u>	0.8	0.7	<u>0.4</u>	<u>2.3</u>	0.9	<u>1.6</u>	<u>1.7</u>
8. Other 16 Groups Combined	0.8	1.0	0.9	0.7	1.0	<u>1.4</u>	<u>1.6</u>	1.1	0.8
Mean for all 27 Groups	1.2	1.0	0.8	0.8	1.1	<u>1.6</u>	1.0	1.1	1.0

* Formula: See text p. 130, Rater and Centre Bias

** Data from one rater only

— = Values < .7 or > 1.3. A value of 1.0 shows no bias

rating of each psychiatrist was compared item by item with the ratings of all the other psychiatrists who assessed the same patient. An Index of Bias was calculated (all those comparisons on which the given rater rated higher, minus all those comparisons on which the rater rated lower than the comparison rater). Ratings were given the values 3 > 2 > 1 > 0 = NP = NA = NI = 0. The results were then grouped into item groups dealing with certain key areas such as hallucinations, delusions, affect, and observed behaviour, and further into "positive interviews" (more items in an interview rated higher than items rated lower compared to other rater) and "negative interviews" (more items rated lower than rated higher compared to other rater). The figures generated by these data can be interpreted as percentages. For example, a bias index of +25 indicates that the investigator rates higher than comparison raters in 25% of the interviews. The statistic does not indicate how much higher he tends to rate, this might range from only a difference of one NR or ? per interview to a much greater discrepancy. These analyses suggested that two of the 26 psychiatrists participating in these ratings tended to rate consistently higher than others regardless of the symptomatology being assessed, while one psychiatrist tended to rate consistently lower. Six raters, three from Aarhus, had a mean Index of Bias across all groups of +25%. The mean bias for all other raters was less than 25%. A review of the raters more recently joining the IPSS showed that they tended to have the greater bias. These data when analysed by centre (Table 8.8) suggested that centre bias also existed in rating areas of pathology and overall level of pathology.

It is not possible in the absence of a definitive method for measuring rater bias to decide between the apparently contradictory results of the two sets of analyses, one showing very little consistent rater or centre bias and the other suggesting that such bias does exist. Probably the best interpretation of these results is that there were in fact consistent biases by centre and by investigator but that these were quantitatively small. The conflicting results might also imply that data analyses for grouping patients and comparing patient samples that dichotomize the values of the PSE data will reflect more bias than those using scaled values.

(7) Agreement on consecutive interviews

A final question concerning the reliability of the PSE, and one of special importance for the use of the PSE in contributing to diagnostic groupings, is how stable the measures of patient characteristics are when evaluated from repeated evaluation interviews. If ratings are to be used for classifying patients into groups, presumably the more repeatable the ratings (at least for brief durations of time, such as a week or so), the more meaningful will be the groupings of the patients derived therefrom as well as the other analyses made from the data. To study this question of repeatability of ratings over time, 51 pairs of consecutive interviews and ratings were carried out by the psychiatrists in the different Centres. In these cases the patient was seen and rated by one psychiatrist and then seen and rated by another psychiatrist within seven days. The agreement of data collected in this way is sometimes considered as "test-retest reliability". However, in this instance the term "reliability" might be misleading since

TABLE 8.8

RATING BIAS IN %* BY CENTRE IN A SAMPLE
OF INTERCENTRE RELIABILITY INTERVIEW*

Field Research Centre	No. of Compar.	Hallucin.	Delusions	Affect	Observed Behaviour	Other Items	Mean	Absolute Mean
HQ	70	-7	-34	-27	-13	9	-12.2	18.0
AARH	209	39	25	3	13	8	17.6	17.6
AGRA	130	23	43	10	7	23	18.5	21.2
CALI	120	-7	18	-22	-32	-14	-10.9	18.6
IBAD	101	22	-54	-4	16	-8	-5.1	20.8
LOND	75	2	15	29	5	-2	10.3	10.6
MOSC	62	-48	-54	-4	13	-3	-19.0	24.4
TAIP	111	-27	25	-5	28	1	4.5	17.2
WASH	52	-55	-19	38	12	8	3.1	26.4
PRAG	82	34	19	10	-41	-5	3.3	21.8
Range		-55,+39	-54,+43	-27,+38	-41,+28	-14,+23	-19,+18.5	10.6,26.4
Total Range		94	97	65	69	37	37.5	15.8

* Formula for Scores: Bias % = $\frac{(\text{number of "+" interviews}) - (\text{number of "-" interviews})}{\text{number of schedules compared}}$

X 100

changes in the status of the patient are being tested together with the reliability of the whole interview process in a test-retest situation. If the values from such analyses are adequately high, they support the use of the method as constructed to define reasonably stable characteristics of patients. Table 8.6 shows the level of agreement for each GUA for all of the pairs of consecutive interviews. It demonstrates that the consecutive interviews have moderately lower values for "reliability" than the simultaneous intracentre ratings, bringing more groups into the level of marginal reliability.

Summary of PSE reliability. The results of PSE reliability evaluations showed generally high reliability by items, UAs and GUAs when the PSE was used by investigators from the same centre. Where investigators from different centres were compared, however, there was less reliability. The lower intercentre reliability appears primarily to have been a result of the different rating situations in the two evaluations. In all situations, ratings made on the basis of observed behaviour had considerably lower reliability than ratings based on patients' responses to specific questions.

8.2.2 Reliability of psychiatric history and social data

The evaluation of whether ratings on the Psychiatric History, Social Description, and other forms have the same meanings within and among centres was far more difficult than the same evaluation for the PSE. On at least two different levels, psychiatric history and demographic events in different cultures have common differences in meaning for identical terms. First, on the descriptive level, the state denoted by "marriage", for example, differs widely among cultures. This problem is reflected in the UN Handbook of Population Census Methods (1959) which states that:

"The apparent uniformity of the major marital status categories enumerated in the various censuses conceals important variations in the definitions of the categories and in many instances a lack of exact definition which is highly damaging to international comparability as well as to the usefulness of results for national analysis."

The same is true for such an important and concrete variable as education. Again, the UN handbook points out that:

"Because of the many differences in national and local school systems throughout the world exact international comparability of census data on level of education completed is probably impossible to achieve. Even the meaning of national figures may be difficult to grasp without a thorough knowledge of the particular educational organization."

With this much difference in the descriptive meaning of relatively simple terms, it is probable that many of the more complex terms like

"personality traits" used in the IPSS forms had widely different descriptive meanings.

The second level of meaning which is far more often a problem with history and demographic data than with symptom data involves the psychiatric implications of a given historical or demographic fact. For example, in some centres the duration of hospitalization reflects the severity of the patient's symptoms. In other centres it is more likely to reflect unavailability of rehabilitative or outpatient facilities, or the absence of a receptive social setting for the ex-patient. Military service at any rank in some countries has a high social value, is eagerly sought after and is evidence of effective personality functioning. In other countries the opposite is sometimes true. Employment history is another factor with complex relationships to local norms.

The interpretations of the results of the Physical and Neurological examinations and laboratory tests (PNE) created similar problems. It became apparent that different terms were used in different centres; and even the same terms had different meanings. On the denotative level, the "abnormal" levels for a sedimentation rate were different in different FRCs (even after correction for the use of different techniques of measurement). On the broader level of significance there were difficulties in equating the psychiatric implications of certain neurological abnormalities between some centres, where they were considered extremely common and of no special psychiatric significance, and other centres where the same abnormalities would have been more often coupled with conditions related to abnormal psychiatric states. It was within the realm of the IPSS to take note of these factors, but future studies will be required to investigate their extent and importance and to develop improved methods for evaluating them.

Despite these problems, a study of agreement between the ratings of psychiatrists from different centres on the same interview was carried out to help evaluate the comparability of history and social data across centres. Although such an exercise does not reproduce the actual process of collecting history and related data in the different centres, it can provide an initial evaluation of the similarity with which representatives of different FRCs rate responses to questions on these subjects. For this purpose, investigators from all Centres were shown the videotape of a history interview and asked to rate it using the standard form.

The results were interesting but sobering. Of 23 coded items for which the majority of ratings were other than "not applicable" or "not inquired", there was perfect agreement among the 15 raters on only 3 items. If only substantive ratings (e.g., presence or absence) are included and ratings of NA and NI are ignored, then 9 items were rated with complete agreement. Of these 9 items all but one was an agreement on absence of a characteristic. Two of the 9 items (divorce, and length of marriage) were actually not applicable since the patient in the history interview had never been married. Aside from these 2, the other 7 items with perfect agreement were: history of obsessions; whether the patient had consulted a professional before coming to the hospital; whether the patient had consulted a non-professional; whether she had a history of alcoholic psychosis; had

a drinking problem, had abused drugs, or had trouble with the law (all rated negative with the exception of having consulted a professional). Besides the 9 items with perfect agreement, there were 4 items with only one or two substantive disagreements out of the 15 raters. These were: history of other hallucinations, aggression, social withdrawal, and history of serious physical illness in a relative living with the patient.

While these 13 items had a high degree of agreement, there were 10 items where at least 3 of the 15 raters disagreed with the majority, or where there was no majority. Strangely enough, these included a wide variety of items ranging from "worked recently" to ratings of recent history of auditory hallucinations, and including ratings of lasting relationships, mental illness in relatives, premorbid personality traits, inpatient and outpatient contacts, history of visual hallucinations, and derealization.

These results indicate that, even in a controlled situation such as this where ratings were made from a technically excellent videotape interview in which the interviewer followed the schedule carefully and the patient was a good and intelligent informant, the reliability of the history ratings may be quite low in many areas. They suggest further that the methods for collecting history data for cross-cultural comparison will need considerable further evaluation and development before they can be used with assurance.

These data are useful for studying ways in which investigators from different centres rate psychiatric history and social information. More work will be needed to investigate the comparability of terms used in reporting facts from the psychiatric history in the different cultures.

While it was difficult to evaluate intercentre reliability on history and social data, it was slightly less difficult to study intracentre reliability. Again, because of the various ways in which these data were gathered in the different FRCs any one paradigm for studying reliability could not provide an adequate test of the reliability of all the data. Nevertheless, to obtain a general estimate of the reliability of the Psychiatric History, 36 patients were rated by pairs of investigators in the centres simultaneously rating the same interview. Of the 36 pairs of ratings on 76 items sampled from the form, there was an agreement on all pairs of ratings for 19 items, agreement on about 90% of pairs for 51 items and agreement on about 75% of the pairs for 4 items (course and duration of present attack, number of outpatient contacts, and number of behavioural symptoms rated present). There was only about 50% agreement on two items (premorbid personality traits prior to age 15, and premorbid personality traits age 15 and over).

The same types of paired simultaneous ratings were made on 32 patients using the Social Description form. Of 36 items sampled there was agreement on all pairs of ratings for 19 items, agreement on about 90% of the pairs for 15 items, and agreement on about 76% of the pairs for two items (socio-economic status of family, and level of current social isolation). As with the PSE, several of the items sampled were always given the same rating for all patients, so that the percentage agreements reported above are somewhat overstated. In general, agreement on history and social data was rather high in these intracentre comparisons. More complex analyses of the data to compensate for invariance were not undertaken at this point, because this

particular exercise could only be an approximate test of the reliability of these forms. Further work is necessary to determine the extent to which psychiatric history and social data can be obtained and recorded in a reliable way.

The third aspect of comparability under consideration in this chapter, reliability of data obtained with repeated evaluations, was not tested for the Psychiatric History and Social Description forms. At this stage the main problems are to develop commonality of meaning and significance of terms among Centres. Until this can be more closely approximated, more complex analyses such as repeated evaluations are premature.

8.3 Agreement on Psychiatric Diagnosis

The assessment of agreement on diagnosis in the IPSS had several aims:

1. To assess to what extent the collaborating psychiatrists meant the same symptom patterns when they used the same diagnostic label.
2. To see which of the characteristics of a patient's psychiatric condition, life history, or social functioning influenced the making of a diagnosis in various Centres.
3. To encourage the collaborating investigators to use the same labels in describing defined symptom patterns even if in their clinical or research practice they use different terms.
4. To familiarize collaborating investigators joining the IPSS at various times with the way in which diagnoses are made for the purposes of the study.

In the course of Phase 2 of the study, videotaped and filmed interviews with patients were used for the assessment of the diagnostic agreement between raters from various FRCs. Prior to this, in the course of Phase 1, joint observation and rating of live interviews had been used in the training of investigators for the study. The joint rating of live interviews was not of course a method that could be widely applied in the study because of language difficulties: only English-speaking patients could be interviewed and assessed by several raters simultaneously because English was the only common language of most of the investigators. Simultaneous translation of interviews was both impractical and expensive. Thus, "live" interviews with several psychiatrists rating the patient were conducted in London and Washington but could only rarely be arranged in other FRCs. The multiple interviews conducted on other occasions were done in the languages of the Centre and the psychiatrists from elsewhere who did not understand the language were rating behaviour only.

At the beginning of Phase 1, but after the training period, a film of a PSE interview was made in London and circulated to all Centres except Moscow. Fifteen collaborating investigators rated the film and all but one agreed on a diagnosis of schizophrenia (ICD number 295), an agreement of

87%. However, when four-digit categories for the subgroups of schizophrenia were taken into account, the level of agreement dropped to 27%. The formula used was:

$$\text{rate of agreement (\%)} = \frac{\text{number of observed agreements}}{\text{total no. of possible agreements}} \times 100$$

Halfway through the study two more patients were diagnosed from videotaped interviews of the PSE. In the case of one patient, the diagnoses were made by 9 psychiatrists from 6 Centres. An overall diagnosis of schizophrenia was made by 8 out of the 9 investigators, showing an agreement of 78% on the 3-digit ICD category 295. Seven out of 9 diagnosed paranoid schizophrenia, the agreement of the 4-digit category thus being 51%. However, for the other patient, there was complete agreement on paranoid schizophrenia between 10 raters from 7 Centres, giving agreements of 100% for both 3-digit and 4-digit categories.

Finally, at the end of the period of data collection for Phase 2, another two patients were diagnosed, again by means of videotaped PSE interviews, by 11 psychiatrists from 7 Centres and from Headquarters. For one patient, 10 psychiatrists agreed on the basic diagnosis of schizophrenia, an agreement ratio of 82%. At the 4-digit category level, 6 were agreed on schizo-affective disorder and 2 on paranoid schizophrenia, while one rated endogenous depression with paranoid traits. For the other patient, the results point up one of the dilemma encountered during the study, namely what to do in the case of a psychiatrist who writes the diagnosis in one category and codes it in a different one. At the 3-digit level, all psychiatrists were agreed on a diagnosis of schizophrenia. If the written diagnosis is taken into account, 8 of the 11 agreed on paranoid schizophrenia and 2 on schizo-affective disorder, giving an overall agreement of 53%. However, if code numbers assigned by the psychiatrist are taken as a basis of comparison, the measure of agreement becomes 65%.

Table 8.9 summarizes the levels of agreement obtained at the three points of time in the study.

It is interesting to note that the psychiatrists made firm diagnoses of their cases in spite of the fact that they were not given a psychiatric history of any other information about the patient except that shown on the videotape of the interview. In order to test whether knowledge of the history causes any change in the diagnosis, a special videotape was prepared that had three parts — a PSE, a history part, and a social description part. The psychiatrists were first shown the videotape recording of the PSE interview. After viewing the tape and rating the schedules they were asked to make a diagnosis. Next they were shown a videotaped recording of the PSE interview. After viewing the tape and rating the schedules they were asked to make a diagnosis. Next they were shown a videotaped psychiatric history and social description interview of the same patient and asked to make a second diagnosis, i.e., either to retain the first one or change it in the light of the additional information. Finally, they were given a narrative account of the patient's history obtained from informants and hospital records and requested to make a final diagnosis (Sartorius et al., 1970).

The patient shown on the videotape was a 41 year old South African who had been living in the United Kingdom for approximately seven years. In

TABLE 8.9

SUMMARY OF PERCENTAGE AGREEMENTS REACHED IN DIAGNOSING 5 PATIENTS
AT DIFFERENT STAGES OF THE IPSS

Time of ratings	Patient No.	No. of raters	Majority Verdict	Percentage level of agreement on ICD	
				3 digits	4 digits
Beginning of Phase 1, but after training	1	15	Schizophrenia, paranoid (4) Schizophrenia, chronic (4)	87	27
Halfway through study	2	9	Schizophrenia, paranoid (7)	78	51
	3	10	Schizophrenia, paranoid (10)	100	100
End of Phase 2	4	11	Schizo-affective disorder (6)	82	21
	5	11	Schizophrenia, paranoid (8)	100	65

the PSE interview he described extremely vivid hallucinations in almost all sensory areas. Connected with these he had a number of delusions. He was well-mannered and gave a very clear account of these experiences. He described the various symptoms that he was experiencing in connexion with the hallucinations and delusions, such as loss of sleep, irritability, and fear. At the time of the interview he realized that his experiences had been abnormal but still was not quite certain whether they were part of a mental illness or caused by his enemies. In the history interview he stated that he had been well until two years ago. He said he was married, well off, had a doctoral degree in pharmacology, and had 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 8.10 shows the three diagnoses each psychiatrist made after the PSE, the history interview, and after additional information had been provided. As can be seen from the table, after they had seen the PSE interview all raters agreed that the patient was schizophrenic. Seven of the 12 agreed on the 4-digit category of paranoid schizophrenia. Having seen the psychiatric history interview these seven raters retained the same diagnosis, and were joined by three others. When given the additional information from informants, two of these three continued to give the diagnosis of paranoid schizophrenia, while the third 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 informants' statements. Another rater added hysteria as a concurrent condition. Six of the 12 psychiatrists maintained the same diagnosis throughout the examination.

It is obvious that most of the psychiatrists used mainly the PSE in arriving at their diagnoses; for some, however, the psychiatric history and other information seemed to influence the psychiatric assessment. This finding, like many others, indicated a possibly fruitful direction of research to be pursued in the followup study.

At the beginning of Phase 2, the psychiatrists were requested to carry out 5 simultaneous and 5 consecutive reliability interviews. After this, they were asked to do one simultaneous interview per month, for each combination of raters in the centers. In all these interviews the psychiatrists were to make their diagnoses independently. The results are shown in Tables 8.11 and 8.12.

Table 8.11 shows the pairs of diagnoses made in 190 simultaneous interviews done in all Centres. Numbers appearing on the diagonal indicate those cases in which there was complete agreement between the psychiatrists. Otherwise the paired diagnoses consist of the interviewer's diagnosis on the horizontal rows and the observer's on the vertical columns. For example, for the 5 cases in the first column, there were 4 in which both interviewer and observer diagnosed simple schizophrenia and 1 where the interviewer's diagnosis was schizo-affective disorder and the observer's was simple schizophrenia.

On 4-digit categories there was complete agreement in 131 cases out of 190, and there were 59 cases in which there was disagreement. Some disagreements are easier to account for than others; for example, it might be

TABLE 8.10 DIAGNOSES MADE BY 12 COLLABORATING INVESTIGATORS ON PATIENT 909 (VIDEOTAPED PRESENT STATE EXAMINATION (PSE) AND PSYCHIATRIC HISTORY (PH) INTERVIEW, AND SUPPLEMENTARY INFORMATION)*

Rater's Identification Number	Diagnosis after PSE only and ICD Number		Diagnosis after PSE and PH and ICD Number		Final Diagnosis after addition of other information	
	ICD Number	Diagnosis	ICD Number	Diagnosis	ICD Number	Diagnosis
012	295.3	Paranoid schizophrenia	295.3	Paranoid schizophrenia	295.3	Paranoid schizophrenia
013	295.9	Schizophrenia, unspecified	781.8	Hallucinatory state	298.9	Reactive psychosis
021	295.3	Paranoid schizophrenia	295.3	Paranoid schizophrenia	292.0	Organic reactive psychosis (? G.P.I.)
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
042	295.9	Schizophrenia, unspecified	295.3	Paranoid schizophrenia	295.3	Paranoid schizophrenia
053	295.3	Paranoid schizophrenia	295.3	Paranoid schizophrenia	295.3	Paranoid schizophrenia
071	295.7	Schizo-affective disorder	295.3	Paranoid schizophrenia	-	-
072	295.3	Paranoid schizophrenia	295.3	Paranoid schizophrenia	295.3	Paranoid schizophrenia
079	295.9	Schizophrenia, unspecified	295.9	Schizophrenia, unspecified	295.3	Paranoid schizophrenia
081	295.3	Paranoid schizophrenia	295.3	Paranoid schizophrenia	301.5	Hysteria
001	295.7	Schizo-affective disorder	295.3	Paranoid schizophrenia	295.3	Paranoid schizophrenia
Agreement on 3 digit category	12 out of 12	Schizophrenia	11 out of 12	Schizophrenia	8 out of 12	Schizophrenia
Agreement on four digits	7	Paranoid schizophrenia	10	Paranoid schizophrenia	8	Paranoid schizophrenia
	8	First and second diagnoses	7	Second and third diagnoses	5	All three diagnoses

* Adapted from Sartorius et al. (1970)

TABLE 8.11 AGREEMENT ON DIAGNOSIS BETWEEN INTERVIEWER AND OBSERVER IN 190 SIMULTANEOUS INTERVIEWS IN ALL CENTRES

Observer \ Interviewer		Schizophrenia 295										Affective		Rest	Rest	Rest	N.D.	Total
		.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	296.1 Rest 296	297	294- 299	300.4	300- 301		
295.0	Schizophrenia, simple	4											1				1	6
295.1	Hebephrenic		8		2				2		1							13
295.2	Catatonic			6		2			2									10
295.3	Paranoid		6		4	1			4					1	1			2 56
295.4	Acute				2	6	1		1		1							2 13
295.5	Latent						2											1 3
295.6	Residual				1						1							2
295.7	Schizo-affective	1		1	1	1			4						1			9
295.8	Other				1				1									2
295.9	Unspecified				2						10						1	13
296.1	Mania								2				6	2		1		11
Rest 296	Other affective psychoses				1				2					13				16
297	Paranoid states														4	1		5
Rest 294-299	Other psychoses														1	12	1	1 15
300.4	Neurotic depression																7	7
Rest 300-301	Personality disorders																1	7 8
	Not diagnosed				1													- 1
	Total	5	14	7	52	10	3	1	18	-	12	6	16	6	15	10	10	5 190

TABLE 8.12 AGREEMENT ON DIAGNOSIS BETWEEN FIRST AND SECOND INTERVIEWERS IN 51 CONSECUTIVE INTERVIEWS IN ALL CENTRES

First interviewer \ Second interviewer		Schizophrenia 295										Affective 296.1	297	Rest 294-299	300.4	Rest 300-301	N.D.	Total	
		.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	Rest 296							
295.0	Schizophrenia, simple	-			1													1	
295.1	hebephrenic		2															2	
295.2	catatonic			2	1							1						4	
295.3	paranoid		2		10								1					13	
295.4	acute					3											1	4	
295.5	latent						-											-	
295.6	residual							-										-	
295.7	schizo-affective								2									2	
295.8	other									1								1	
295.9	unspecified										2							2	
296.1	Mania											-						-	
Rest 296	Other affective psychoses								1			2	4					7	
297	Paranoid states												1	2				3	
Rest 294-299	Other psychoses								1						1			2	
300.4	Neurotic depression												2					2	
Rest 300-301	Other neuroses and personality disorders							1									3	4	
	Not diagnosed									3			1					4	
	Total	-	4	2	12	3	-	1	4	4	2	2	9	3	1	-	3	1	51

likely that if one rater diagnosed schizo-affective disorder, the other might diagnose either a form of schizophrenia or depression. There were 19 disagreements in which one psychiatrist diagnosed schizo-affective disorder: in 14 of these the other diagnosed another form of schizophrenia, in 4 affective disorder, and in only 1 another type of psychosis.

However, there were cases in which the disagreement might be regarded as more serious, one diagnosis being a psychosis and the other a neurosis. This occurred in only 5 cases of the 190. The two most interesting of these concerned one case where the diagnoses were mania and neurotic depression and one case where a diagnosis of neurotic depression was combined with that of reactive depressive psychosis.

Table 8.12 shows the distribution of pairs of diagnoses made in 51 consecutive interviews. In this type of interview one would expect a higher level of agreement in the diagnoses than in the symptomatology on which these are based, since some symptoms might be modified between interviews even if those on which the diagnoses are founded would be unlikely to disappear. In these patients there was diagnostic agreement in 32 out of 51, or 39%, compared with 47% in the simultaneous interviews. If the psychiatrists were perfectly consistent from one interview to the next, the difference between the first and second consecutive interviews would be due to changes in the patient's condition. In this sample the increase of 8% in disagreement in the consecutive interviews is in very good agreement with the 5% cited from Ward et al. (1962) in Chapter 2.

8.4 Summary

From the assessment of applicability and reliability of the methods, four major conclusions can be drawn.

1. The PSE interview can be administered satisfactorily in various cultures and in widely different psychiatric centres. This semi-structured type of mental status interview fitted easily in both form and content into the assessment techniques already used by investigators in the Centres and was applicable to the patients seen.

2. It is possible to obtain high reliability among interviewers within centres using the PSE. However, the level of reliability across centres, while high enough to encourage cross-cultural comparisons of data, implies that further work on inter-centre reliability is necessary. Improvements will have to be made both in methodology for studying cross-cultural reliability and possibly in construction of schedules and in rater training to increase intercentre reliability to a higher level. Items with the lowest levels of reliability are those based on observation rather than on patient reports and include behaviour, affect, and rapport. Since these measures are often most crucial for the diagnosis of catatonic, hebephrenic, and simple schizophrenia, comparisons of these subgroups among Centres are especially speculative. A more specific statement of how the reliability of specific items affects confidence in diagnostic comparisons will be presented in Chapters 10 and 11 dealing with diagnostic methods. The question of whether

the marginal level of intercentre reliability for many areas of symptomatology is also related to consistent rating bias between centres remains unanswered, since one method for evaluating bias finds only minimal bias by centre while the other suggests that significant rater and centre biases are both present. Whatever the sources of the lower reliability of intercentre ratings, in future studies of this kind it will be worthwhile to pay even greater attention to increasing the reliability of ratings so as to highlight further the similarities and differences among patients from different centres.

3. Psychiatric history and demographic data, and to some extent psychiatrically pertinent physical data, are much more difficult to obtain in a comparable way cross-culturally. This difficulty stems from the complexity of the data to be collected, the variety of sources of such data, and the problems of relating social factors in one culture to those of another culture. Nevertheless, for purposes of comparing the patients' background and social functioning, in order to relate these data to psychiatric symptomatology, and for purposes of adequate follow-up and morbidity ratings in different cultures, the methodology of collecting and analyzing such data is an extremely important area to be developed further.

4. Intercentre reliability in making diagnoses could only be tested on multiple ratings of videotaped interviews. It appeared satisfactory so far as 3-digit categories of the ICD were concerned but less so when the 4-figure diagnoses were compared. A similar conclusion emerged from a study of the diagnoses made on the basis of simultaneous and consecutive interviews carried out in each Centre.