CLASSIFICATION OF THE BEHAVIOR DISORDERS

By JOSEPH ZUBIN

Biometrics Research, New York State Department of Mental Hygiene, and
Columbia University, New York, New York

An essential step for clarifying a field that is as confused as the classification of behavior disorders is to provide some specific definitions for the terms to be used. While definitions can never be rigorous and complete except in mathematics, they nevertheless serve to demarcate a concept even though its boundaries remain somewhat blurred. Classification, following Hempel (19), divides a set of objects into subclasses according to certain common characteristics. In the field of behavior disorders, a classification system attempts to classify the population of those suffering from behavior disorders into diagnostic subgroups according to certain common signs, symptoms, and other identifying characteristics. The resulting classification system will depend upon several basic considerations. First, is the goal of the classification, second, the definition of the universe of patients, and third, the specification of the various subcategories into which the population under examination is classified. Variations in these goals and definitions will result in placing the same individual in different categories or in an excluded category outside of the universe of discourse. For this reason it is important to define the goals, the universe of discourse, and the subcategories according to which classification is made.

The main purpose of diagnosis is threefold: aetiology, prognosis, and selection of therapy. The search for aetiology consists of finding the group of possible causative factors accounting for the symptoms and signs and then the diseases which are associated with these causes. From this group of diseases the most likely and the least likely are separated and further testing initiated to determine which disease is the most probable. Once the diagnosis is arrived at, the clinical data are used to estimate how the patients who fall into the given disease category will fare under the variety of treatments available and to discover to which subgroup the patient in question belongs, so as to select for him the most suitable therapy.

Another use is made of the clinical data. Not all the signs and symptoms observed by the clinician constitute a part of the syndrome basic to the disorder. Some of these signs and symptoms, distressing though they

1 The author wishes to acknowledge the cooperation of Mr. Percy Tilton in organizing the plan for this review, supervising the collection and abstracting of the articles referred to, and in the editorial work involved in bringing this chapter to publication. He was ably assisted in this work by Miss Jane Cadwallader and Miss Vivian Hoff. The author is grateful to Mr. David Jenness for reading the manuscript and making valuable suggestions. This review was supported in part by USPHS Grant No. MH-09191.
are to the individual, are purely coincidental to the basic disorder. The
clinician, however, must deal with them in addition to dealing with the dis-
order itself. It is quite possible that because of our ignorance, most of the
therapies in the field of behavior disorders may be dealing only with the
incidental aspects of the patient’s distress and may not deal at all with the
basic illness nor with its prognosis. Thus, some of the symptoms and signs
which we regard as our chief targets may postpone needlessly the approach
to the chief source of the disorder. It is clear that, at the present time, the
specific goals of diagnosis must be spelled out if we are ever to reach any
agreement in regard to aetiology and prognosis. It is also quite clear that
the goals of the clinician, epidemiologist, phenomenologist, etc. can be
quite disparate and it is this disparity which may be leading to confusion
and bringing diagnosis into disrepute.

The multilevel purpose of diagnosis may be described as follows: the
immediate purpose of diagnosis is clinical—namely, finding a suitable treat-
ment. This practical immediate goal is pragmatic and need not wait neces-
sarily for the other purposes of diagnosis. We use aspirin pragmatically
for headaches even though the scientific basis for its action is unknown,
the cause of the headache not determined, and the eventual outcome of the
ailment not discerned. One aspect of diagnosis dealing with selection of
treatment in the mental disorders may be likened to the work of the prac-
tical green-thumb gardener rather than to that of the scientific botanist. But
we must not confuse the full picture of what the ideal diagnosis should
consist of with the limited aspect provided by the pragmatic approach. For
example, Stockings (62) has suggested that we might substitute for the
current diagnoses of the schizophrenias a system which would indicate the
most suitable treatment for the case in question. He made this suggestion
during the period when the shock therapies were popular and suggested the
following classifications: patients who could be successfully treated with
insulin, whom he designated as dysglycolytic, and electroconvulsive ther-
apy (ECT) candidates, whom he designated as dysoxic. One could continue
this classification system by specifying patients who are most successfully
treated with phenothiazine, with psychotherapy, etc. The rapid change in
the therapeutic armamentarium, however, would render such classification
systems obsolete, but the basic idea of classifying patients according to the
most suitable therapy of the day is a good example of how one aspect of
diagnosis could be made the focus of the entire process. If we succeed in
finding, for example, the most suitable therapy for the various types of pa-
tients that we now can discriminate, be it behavior therapy, drug therapy,
group therapy, psychotherapy, or what not, we may make more progress in
the immediate future than by searching for aetiology and prognosis inde-
dependently of the therapeutic aspect. That this path involves dangers is il-
lustrated by the fate of the ear-nose-and-throat specialty in physical medi-
cine. The arrival of antibiotics has practically wiped out the specialty and
with it a field of aetiology and prognosis which might have added much to
knowledge. Will the drugs or behavior therapies similarly undermine the clinical diagnostician in the field of mental disorders? The likelihood is that they will not, since unlike penicillin our therapies are not yet cures but merely mitigation of the mental disorders.

By behavior disorders we mean those types of behavior which deviate from expected social-cultural norms and which are deleterious either to the person or to society or to both. These disorders tend to be either progressive or stationary, but have in common the tendency to shorten life or severely reduce efficiency and happiness or both. The stationary disorders differ from the progressive only in so far as they do not necessarily shorten life, but they do interfere seriously with efficiency and happiness. This definition of behavior disorders is sufficiently general to encompass all known disorders in most nosological systems. It is in fact too broad; it includes criminal behavior, which we shall try to exclude from our consideration. The definition differs from most others in that it sidesteps structural and aetiological considerations. For many of the behavior disorders very little is known about these two factors. In fact, when the aetiology or structural defect accompanying a behavior disorder becomes known, the disorder is usually lost to some other discipline as was the case with general paresis, phenylketonuria, etc. The field of behavior disorders, as we view it, retains for the most part only the disorders of unknown origin.

It should also be noted that behavior disorders assume a value system in which the presence of the disorder is regarded as undesirable. Since this value system is culturally determined, it stands to reason that some behaviors will be regarded as disorders in some cultures but not in others. The dilemma of defining disorders whose sole condition is deviant behavior has been strikingly portrayed by Wootton (75):

the anti-social behaviour ... is the precipitating factor that leads to mental treatment. But at the same time the fact of the illness is itself inferred from this behaviour: indeed it is almost true to say that the illness is the behaviour for which it is also the excuse. But any disease, the morbidity of which is established only by the social failure that it involves, must rank as fundamentally different from those of which the symptoms are independent of social norms.

This distinction will, moreover, still remain even if we reach the stage, as we very well may, when every mental process has its known physical accompaniment, and when our present dualistic language, along with the distinction between 'organic' and 'functional' disorders, can be discarded.

Regarding the population of individuals to which the categories apply, we shall limit ourselves to those seeking help for mental disorders either in a hospital, clinic, or office of a private practitioner. Although this definition includes many who are not suffering from behavior disorders and excludes many who do, it is nevertheless a good working definition since most of the literature, excepting that concerned with surveys of the general population, deals with populations included in this definition.

The present status of the classification of behavior disorders is, to say
the least, chaotic. There are at least 50 different types of classifications in varying degrees of use throughout the world ranging from those which deny the existence of behavior disorders as entities (64, 65, 66) to those who regard all behavioral disorders as manifestations of a single underlying dimension—incapacity to cope with life's vicissitudes (42, 43, 44)—to those who believe that there are differentiable entities in the field of behavior disorders just as there are entities in the field of physical disorders.

By diagnoses are meant the subclasses into which a classification system distributes the members of a given population. The individuals belonging in a given subclass possess certain characteristics which all of them and they alone have in common. "Each subclass is defined by . . . a certain concept which represents the complex characteristics essential for membership in that subclass" (19).

The concept underlying a given diagnosis may be derived from mere description of patient characteristics or may have either a prognostic, structural, aetiological, or therapeutic implication. In so far as it is based on mere description alone without any theoretical import it may prove to be an artificial category and not one which so to speak "carves nature at the joint" because its defining characteristics have many explanatory connections with other traits (19). The search for meaningful underlying concepts has been in progress for many decades and some of the historically significant concepts are: (a) Freud's unconscious, (b) Janet's energy level, (c) Bleuler's loosening of association, (d) Selye's concept of stress, (e) Jung's introversion-extroversion; but they have not provided theoretical models by which bridges could be built to the still unknown areas of psychopathology. Until such concepts are provided we may have to be satisfied with descriptive methods that do not have a longer explanatory aetiological reach. But for progress, a gradual development of fruitful conceptual frameworks is essential.

While waiting for the arrival of the necessary explanatory framework we might take a look at the present state of our classification system with regard to its reliability and validity. Following a review of these two areas, a review of the newly emerging trends will be undertaken.

**Previous Reviews**

In his 1950 chapter, Hunt (22) concluded that the focus of attention during the period of his review (1949) was on the subjective clinical interpretation of responses obtained from projective and cognitive techniques rather than on objective quantifiable evaluations of behavior. Furthermore, the foremost problem of the day was that of the validity of the behavioral evaluations, and though psychiatric diagnosis offered a ready criterion, such diagnosis was considered unacceptable because of its low reliability.

Lorr, in his review (39), stressed the difference between the typological approach of Kraepelin in which specific empirical syndromes of observable behavior characterized the diagnostic categories and the more recent
configural approach in which not all the elements of the syndrome needed to be present in order for a patient to belong to a given diagnostic group. The diagnosis is not merely a question of the presence or absence of a specified number of symptoms which constitute the disorder, but is instead an ordering of each of the elements (signs and symptoms) of a given disorder on a continuum of intensity. In this way Sheldon transmuted Kretschmer's body type categories into subclasses determined on the basis of a set of three dimensions (ectomorphy, mesomorphy, and endomorphy). This distinction between the typological approach and the dimensional approach raises some interesting problems in the diagnosis of the behavior disorders. Types thrive on the presence of discontinuities either in a given dimension or in the relationship between dimensions. Whether these discontinuities inhere in the behavior of the patient or in the behavior of the observer (his judgment of the presence of deviance) is still an open question. If the behavior of the patient group is not separated by a discontinuity from the behavior of the rest of the population, but instead it is the observer who creates the discontinuity, the relativity of diagnosis in different cultural groups is a necessary result (84).

The ten years between Hunt's and Lorr's reviews brought major changes of emphasis—not in the basic theoretical issues but in the methods of solution. Lorr's review reflects a preoccupation of many research workers with establishing observable quantitative symptomatology rather than subjective clinical tests as the criteria of diagnostic differentiation. In essence, however, the central problem in the field of the classification of the behavior disorders remains the same, namely, the reliability and validity of differential diagnostic conceptualizations and criteria.

The intent of the present review is to examine the literature for the period from January 1960 through December 1965 for selected studies which are directly or indirectly concerned with establishing the reliability and validity of our current diagnostic system of psychiatric classification. Where pertinent, earlier publications will be cited. The following journals have been thoroughly searched for relevant material: American Journal of Psychiatry, Archives of General Psychiatry, Comprehensive Psychiatry, British Journal of Psychiatry, British Journal of Social and Clinical Psychology, Canadian Psychiatric Association Journal, International Journal of Psychiatry, Journal of Abnormal and Social Psychology, Journal of Psychiatric Research, Psychiatric Research Reports, Psychiatry, and the Journal of Nervous and Mental Disease. Additional sources have been scanned.

Among the outstanding books which have appeared during the period of this review are the following: (a) Jaspers' General Psychopathology (25), which, translated into English from the 1923 and 1946 editions, brought to the attention of the English-speaking world his classic approach to psychopathology; (b) Menninger, Mayman & Pruyser's The Vital Balance (44); (c) Engel's Psychological Development in Health and Disease (15); (d)
Lorr, Klett & McNair's * Syndromes of Psychosis (40); (e) the series of field studies by the Cornell group: *Mental Health in the Metropolis. The Midtown Study (60), Life Stress in Mental Health (34), My Name is Legion (35), People of Cove and Woodlot (21), The Character of Danger (36), Approaches to Cross-Cultural Psychiatry (46); (f) Hoch & Zubin's edited volume *Comparative Epidemiology of Mental Disorders (20); and (g) Zubin's edited volume, *Field Studies in the Mental Disorders (81). Two recently held and still unpublished conferences on the problem of diagnosis are the *Conference on the Role and Methodology of Classification in Psychiatry and Psychopathology under the auspices of the American Psychiatric Association and the Psychopharmacology Service Center of the NIMH (Nov., 1965) (84), and the continuing *Symposium on Definition and Measurement of Mental Health (87) under the auspices of the National Center for Health Statistics, USPHS. To review these books alone would require an entire chapter but pertinent references to these volumes will be found in this review.

Historically, mental disorders were first classified according to individual symptoms. This gave rise to a multiplicity of diagnoses, which numbered as high as 2400 (44). It soon became necessary to subgroup them into "individual connections, symptom-complexes, causal relations, etc., until the idea of the disease-entity came to have a significance of its own for diagnostics—a significance which can never be final" (25). After pointing out the difficulties in the way of developing a comprehensive schema of psychoses, Jaspers explains why, despite the impossibility of achieving such an end, the struggle still persists.

In the first place we want to see properly what this idea of disease-entity has achieved in respect of the over-all picture of existing psychic disorders, and particularly where we have failed because it is the basic and radical failures which make us aware of the actual state of our knowledge. In the second place every presentation of special psychiatry requires some classification of psychosis as its base. Without some such schema it cannot order its material. In the third place we need a classification in order to make statistical investigations of a large case material.

According to Jaspers there are three types of psychiatric disease concepts. The first of these is based upon the idea of disease as possessing some "somatic process as an essential part of the illness" which may differentiate and define it—"In fact there is a field of organic cerebral disease where the demand for a somatic basis can be gratified and where the psychic events are symptoms of a known physical event." The second type of disease concept is completely dependent upon psychic changes which take place within the individual.

When investigating these illnesses we would like to discover—as with the first group—'basic functions' in the psychic events, the disturbances of which could make the manifold phenomena comprehensible. We would not uncover, it is true,
the somatic process but we would find what was the specific factor and particularly in the case of schizophrenia what was the new element in contrast to the healthy state. By purely psychological means we could discover something of the nature of the illness though the assistance of theoretical concepts would still be needed.

Jaspers’ third type of psychiatric disease concept is that of extreme deviations from so-called normal behavior.

The illness—in spite of the gap between health and the neurotic mechanism—does not bring anything new in principle as contrasted with the previous state of health although developments can be initiated which prove ruinous for the psyche. There are basic properties of human existence which show themselves in the exceptional case more markedly, effectively and alarmingly than in the majority of people.

Jaspers, then, proceeds to develop three major groupings:

I. Known somatic illnesses with psychic disturbances
   1. Cerebral illnesses: traumas, tumors, acute and chronic infections, vascular diseases, hereditary atrophic system diseases (Huntington’s, Pick’s, Parkinson’s), organic deterioration associated with age
   2. Systemic diseases with symptomatic psychoses: infections; endocrine disorders; etc.
   3. Poisons: alcohol, drugs, carbon monoxide

II. The three major psychoses
   1. Epilepsy
   2. Schizophrenia
   3. Manic depressive illnesses

III. Personality-Disorders (Psychopathien)
   1. Isolated abnormal reactions that do not arise on the basis of illnesses belonging to Groups I and II
   2. Neuroses and neurotic syndromes
   3. Abnormal personalities and their developments

Only the disorders in Group I satisfy the ideal requirements of disease-entities ontologically since only for them is the site of the disorder, the malfunction, and the aetiology known. But only a small proportion of mental patients fall into Group I and, furthermore, most of them are treated by disciplines other than clinical psychology and psychiatry. The bulk of the hospitalized disorders are in Group II and the unhospitalized ambulatory disorders in Group III. It should be pointed out that the epilepsies with their convulsive features and electroencephalographic (EEG) manifestations are closer in some respects to Group I than to Group II since cases in the latter have no invariant behavioral characteristic by which they can be recognized. Group III is the least reliably differentiated since it is made up of extreme variations of personality which do not show any abrupt discontinuities from the normal.

It follows that diagnosis proper is only possible and necessary in Group I. With
Group II the majority of cases will fall by consensus of contemporary psychiatric opinion into one of the three major psychoses but the diagnosis has got no specific alternative character. Either it is clear as a whole or the differential diagnostic discussion over details determines nothing. In Group III the only thing which is of value is an extensive analysis of the case, in its phenomenological, meaningful and causal aspects and a precise grasp on the personality, its reactions and life-history and its more weighty developments; but apart from some separation into a great number of type-groupings, diagnosis is an impossibility.

Hence in Group I our diagnosis is according to the classes of diseases to which a case either does or does not belong. In Group III it is according to types many of which can be found together in the same case according to the particular point of view. In Group II we have classes of disease in mind although their definitive causes and nature are not known, but in fact one is always confined to types.

Jaspers, being very cognizant of the limitations inherent within conceptions about diagnostic systems which rely upon acceptance of ideas about psychiatric disease-entities, prefaces the development of his own schema in the following manner:

When we design a diagnostic schema, therefore, we can only do so if we forego something at the outset. We abandon the idea of disease-entity and once more have to bear in mind continually the various points of view (as to causes, psychological structure, anatomical findings, course of illness and outcome) and in face of the facts we have to draw the line where none exists. Such a classification therefore has only a provisional value. It is a fiction which will discharge its function if it proves to be most apt for the time. There is no ‘natural’ schema which would accommodate every case. Even the most experienced psychiatrist comes across a number of cases repeatedly which are new to him and which he cannot classify no matter what schema he applies. . . .

**RELIABILITY**

Although the concept of reliability is well-known and frequently applied in psychometrics, it is a relative newcomer in clinical diagnosis. The well-known method of odd-even reliability is, of course, not applicable to diagnosis. There are, however, at least two types of reliability which are quite applicable: (a) agreement between two observers, and (b) agreement between an initial and a subsequent diagnosis. The latter we shall designate as agreement, the latter we shall designate as consistency. There is still a third method which compares the frequency of each diagnosis in two random samples of the same population. If the proportion of patients in the various diagnostic categories agrees in the two samples, there is presumptive evidence of high level of agreement in diagnosis between the different observers. We shall designate this measure of reliability as “frequency agreement.”

*Observer agreement.*—In studies of agreement between different observers there was of course considerable variation in the types of clinicians participating, types of institutions, types of patient utilized, and manner of conducting the interviews. To facilitate levels of comparison between in-
vestigations, the author has at times reworked reported data to make the
measures of agreement more comparable. The degrees of agreement are, in
general, of two kinds:

(a) Criterion Agreement. Here official diagnoses by a senior judge or a
group of senior judges is the criterion by which other judgments are evalu-
ated; (b) Average Group Agreement. This may be obtained by one of two
methods when no criterion judgment is available. These methods regard all
the judges as being equal in judgmental capacity, and the average agree-
ment between all possible pairs of judges is taken as the degree of agree-
ment. Thus, if only two judges are available, the percentage agreement be-
tween them is the measure sought. If there are \( n \) judges each rating \( m \)
cases, there are \( m(n) (n - 1)/2 \) pairs of judgments and the proportion of
these which are in agreement with regard to a given category is the measure
of agreement sought. Thus the numerator consists of all concordant diagnoses
with regard to the presence of a given disorder or its absence and the de-
nominator consists of all the possible pairs of judgments both concordant
and discordant. This may be regarded as the average agreement between all
pairs of judges. The second method of obtaining average group agreement
differs from the first only in that for each diagnostic category—whether a
general or specific category—the degree of agreement is obtained by collect-
ing into one group only the patients who were diagnosed at least once as
belonging to a given category. Then, the proportion of concordant diagnoses
for the category in question is related to the total number of diagnoses made
for the patients in question.

Data on degree of agreement are presented in Table I.

The percentages of agreement for the more specific Organic categories
are far from similar: for acute brain syndrome, the values are 46 per cent
and 68 per cent; for chronic brain syndrome, 66 per cent and 80 per cent;
for mental deficiency, 42 per cent and 73 per cent.

For the specific Functional Psychoses wide differences are found: for
schizophrenia the degree of agreement ranges from 53 per cent to 80 per
cent (not including Kreitman's 0 agreement for six cases); for the general
category of affective psychoses, the figures are 35 per cent and 65 per cent;
for involuntional psychoses, from 26 per cent to 57 per cent; for manic de-
pressive psychosis (depressed), 36 per cent and 82 per cent.

For the categories of Psychoneurosis the degree of agreement appears
to be low and varied. For neurosis (unspecified), from 16 per cent to 56
per cent; reactive depression, 18 per cent and 63 per cent; for anxiety
state, 27 per cent and 55 per cent.

For the Characterological categories (personality pattern disturbance,
personality disorder, personality trait disorder, sociopathic disturbance),
the agreement is also low and variable with the exception of sociopathic,
where the values are 54 per cent and 58 per cent. However, the range
across the other categories in this grouping is large, from 6 per cent to 66
per cent.
### TABLE I
PERCENTAGE AGREEMENT ON DIAGNOSIS BY OBSERVERS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DESCRIPTION:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location of study</td>
<td>US</td>
<td>US</td>
<td>US</td>
<td>GB</td>
<td>US</td>
<td>US</td>
</tr>
<tr>
<td>Med. facility</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>N of patients</td>
<td>52</td>
<td>6</td>
<td>426</td>
<td>90</td>
<td>153</td>
<td>91</td>
</tr>
<tr>
<td>Sex</td>
<td>M</td>
<td>—</td>
<td>MF</td>
<td>MF</td>
<td>MF</td>
<td>MF</td>
</tr>
<tr>
<td>Class. type</td>
<td>X(1, 2)</td>
<td>—</td>
<td>X(1)</td>
<td>Y</td>
<td>X</td>
<td>X(1)</td>
</tr>
<tr>
<td>Interval</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Agreement</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td><strong>GENERAL CATEGORIES:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organic</td>
<td>92</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional psychosis</td>
<td>80</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Characterological</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>74*</td>
</tr>
<tr>
<td>Psychoneurosis</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. agreement, gen. cat.</td>
<td>64††</td>
<td>84</td>
<td>78</td>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ORGANIC PSYCHOSES:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute brain synd.</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46</td>
</tr>
<tr>
<td>Chronic brain synd.</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>Mental deficiency</td>
<td>(42)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>73</td>
</tr>
<tr>
<td>Psychosis of old age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>Paralysis</td>
<td>49†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>62†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FUNCTIONAL PSYCHOSES:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>80†</td>
<td>75††</td>
<td>(0)</td>
<td>53</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Affective</td>
<td>(35)</td>
<td>65†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Involutional</td>
<td>57</td>
<td></td>
<td>40</td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>Manic depressive (depressive)</td>
<td>82†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>Psychotic depressive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>Hypomanic</td>
<td>29†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paranoid reaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Other reaction</td>
<td>(16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(17)</td>
</tr>
<tr>
<td><strong>NEUROSIS AND PERSONALITY DISORDERS:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychoneurosis</td>
<td>(16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>Reactive depressive</td>
<td></td>
<td></td>
<td>18†</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety state</td>
<td>(8)</td>
<td></td>
<td>27†</td>
<td>55</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>Psychophysiological reaction</td>
<td>(6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>Pers. pattern disturb.</td>
<td>(8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pers. disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>Pers. trait disturb.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociopath</td>
<td>58</td>
<td></td>
<td>54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AV. AGREEMENT ON SPECIFIC CATEGORY</strong></td>
<td>38†#</td>
<td>66†</td>
<td>55</td>
<td>63</td>
<td>54</td>
<td>57</td>
</tr>
</tbody>
</table>

---

**LEGEND TO TABLE I**

* Type of medical facility:
  1 = Clinic
  2 = Psychiatric hospital
  3 = Psychiatric unit of general hospital

* Type of Classification:
  X = APA Diagnostic & Statistical Manual, 1952
  (1) abbreviated
  (2) with additions
  Y = Special

* Interval between interview:
  0 = Simultaneous (or within a few minutes)
  1 = 1 week
  2 = 2 weeks
  3 = 1-2 months

* Kinds of agreement:
  A = Criterion
  B = Average group

---

**SYMBOLS:**

--- = Unspecified

( ) = N < 8

† = Calculated on data reported in another form

‡ = Criterion agreement, reworked by present writer. Schmidt & Fonda’s 51% agreement refers to specific subtypes of schizophrenia

# = 3 psychiatrists, 2 at a time

* = The combined av. of character disorders and psychoneuroses
Agreement in regard to unspecified general Organic diagnoses is rather high, 85 per cent and 92 per cent; for the Functional Psychoses, from 71 per cent to 80 per cent; approximately 72 per cent for the Characterological Disorders; and in the one reported study for general Psychoneurosis the agreement is 52 per cent. The overall level of agreement for broadly viewed diagnostic categories ranges from 64 per cent to 84 per cent.

A review of this table indicates that the degree of overall agreement between different observers with regard to specific diagnoses is too low for individual diagnosis. The overall agreement on general categories of diagnosis, although somewhat higher, still leaves much to be desired. The evidence for low agreement across specific diagnostic categories is all the more surprising since, for the most part, the observers in any one study were usually quite similar in orientation, training, and background.

A more detailed analysis of the studies in Table I reveals some interesting trends not describable in the tabulations given there. Only one of these studies, that of Ash (2), concludes with a despairing note regarding reliability, but the low reliability can be partially explained on the ground that this study dealt with an outpatient group consisting of patients with minor ailments. All of the other reporters, though admitting difficulties, consider the present diagnostic procedures quite reliable. Schmidt & Fonda (54) give the impression of having come to scoff and stayed to pray, but their rosete conclusions are somewhat inflated because their reliability for specific diagnoses often depends upon fourfold tables in which the category under examination is contrasted with all other categories. As an example they themselves point out that in their Table 3 (p. 266) the agreement between resident and chief psychiatrist is inflated when schizophrenia is contrasted with all other categories. Despite the fact that the nonschizophrenia category might show considerable discrepancy with regard to specific diagnosis between resident and chief, all the patients in this nonschizophrenia category are regarded as concordant for the purposes of computation. In their Table 5 they present the proportion of the chief's diagnoses, which were concordant with the residents' diagnoses when only psychotic patients were considered; but this too is a little high, since the diagnoses of schizophrenia made by the resident which were not concurred on by the chief are excluded. Referring to the authors' Table 3, a better measure is to include in the denominator all the patients who were given a diagnosis of schizophrenia by at least one diagnostican, and in the numerator all the patients on whom agreement was reached. The proportion would become 99/(99 + 19 + 21) = .71. Unfortunately, their Table 5 does not give the data with which to compute this type of agreement for the other categories and hence the proportions given, small as they are, are in all probability too high. It is also interesting to note that although the agreement between the residents and the criterion (chief's diagnosis) is .75 for the total category of schizophrenia, for the subcategories it is only .51.

Kreitman et al. (32) point out that the purpose of diagnosis often de-
terminates the degree of reliability that is needed, the need for high reliabil-
ity being, for example, greater in epidemiological studies than in clinical
work, where validity is far more important than high reliability. In exam-
ining background variables in relation to diagnostic reliability, they found
that agreement between judges on the presence or absence of previous
mental illness was associated with concordance on diagnosis and that the
higher frequency of previous illness also was associated with concordance
on diagnosis. The other background variables (availability of informants
and whether they were the same or not on two occasions; duration of illness;
family history and whether this was positive or negative for mental illness)
did not influence the concordance in diagnosis. An interesting suggestion is
made by the authors with regard to the role of severity of illness. It ap-
pears that the more severe the illness, the greater the concordance. How-
ever, mere agreement in diagnosis is not enough to establish the value of
the diagnoses unless the basis for the agreement is specified. Nor, on the
other hand, is reporting of disagreement enough, since the basis for the
disagreement must be specified if progress is to be made.

Consistency of diagnosis over time.—The stability of diagnosis has
often been taken as another indicator of the reliability of diagnostic cat-
ergories. One of the first studies to find that consistency was low is that of
Wilson & Deming (72) describing the change in diagnosis that occurred
when patients were moved from the Boston Psychopathic Hospital to vari-
ous state hospitals. It is, of course, important to realize that disagreement
of an initial and a subsequent diagnosis may reflect changes in the patient's
symptomatology rather than lack of consistency. Very often, even a few
days in the hospital will alter a patient's behavior sufficiently to alter his
diagnosis. Moreover, the availability of psychopharmacological agents dur-
ing this interval of time may have dramatic effects on patient behavior and
in turn on the assigned diagnosis. For these reasons, it is important to sepa-
rate studies in this area into those with a short interval between examina-
tions (several weeks) and those with longer intervals (greater than four
weeks). One of the most persistent attempts at investigating the consistency
of diagnoses was made by W. A. Hunt and his associates on data provided
by the neuropsychiatry screening program of the U. S. Navy during the sec-
ond World War (23). Hunt viewed the diagnostic procedure as an example
of decision making and investigated the reliability of such judgments ranging
from the overall view of detecting the presence of general psychopathology
(unfitness for service) to the more specific view regarding the exact nature
of the psychopathology. Since the Navy provides rather specific criteria of
suitability or unsuitability for service, he examined the degree of agreement
between initial and subsequent judgment of suitability for service and found
it to be quite high—93.7 per cent. However, with regard to the major diag-
nostic categories, the degree of agreement dropped to 54.1 per cent; and for
specific diagnoses to 32.6 per cent. The data are shown in Table II.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DESCRIPTION:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location of study</td>
<td>U.S.</td>
<td>G.B.</td>
<td>Norway</td>
<td>U.S.</td>
<td>U.S.</td>
</tr>
<tr>
<td>1st med. facility</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1, 2, 5, 8</td>
</tr>
<tr>
<td>2nd med. facility</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>1, 2, 5, 8</td>
</tr>
<tr>
<td>N of patients</td>
<td>794</td>
<td>6263</td>
<td>200</td>
<td>218</td>
<td>4512</td>
</tr>
<tr>
<td>Type of patient</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>C, D, B</td>
</tr>
<tr>
<td>Class. type</td>
<td>Y</td>
<td>W</td>
<td>V</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Interval</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Time of diagnosis</td>
<td>ad.</td>
<td>ad.</td>
<td>ad.</td>
<td>dl.</td>
<td>ad.</td>
</tr>
<tr>
<td>Independence of diagnosis</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>ORGANIC PSYCHOSES:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic brain synd.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental deficiency</td>
<td>70*</td>
<td>89*</td>
<td></td>
<td>10*</td>
<td></td>
</tr>
<tr>
<td>General paralysis</td>
<td>53*</td>
<td>92*</td>
<td></td>
<td>53*</td>
<td></td>
</tr>
<tr>
<td>Senile dementia</td>
<td>29*</td>
<td>63*</td>
<td></td>
<td>92*</td>
<td></td>
</tr>
<tr>
<td>Cerebrovascular</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epileptic</td>
<td>72*</td>
<td>92*</td>
<td></td>
<td>92*</td>
<td></td>
</tr>
<tr>
<td>Alcoholic</td>
<td>64*</td>
<td>53*</td>
<td></td>
<td>53*</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>60*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total organic</td>
<td>53*</td>
<td></td>
<td></td>
<td></td>
<td>65*†</td>
</tr>
<tr>
<td><strong>FUNCTIONAL PSYCHOSES:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>55*</td>
</tr>
<tr>
<td>Schizophreniform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28*</td>
</tr>
<tr>
<td>Schiz., schizo-affective</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear schizophrenic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total schizophrenic</td>
<td>37*</td>
<td>68*</td>
<td>87*</td>
<td>66*</td>
<td>70*</td>
</tr>
<tr>
<td>Involutional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10*</td>
</tr>
<tr>
<td>Manic depressives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10*</td>
</tr>
<tr>
<td>Affective</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10*</td>
</tr>
<tr>
<td>Paranoid</td>
<td>29*</td>
<td>31*†</td>
<td></td>
<td>43*</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total functional</td>
<td>59</td>
<td></td>
<td></td>
<td>43*†</td>
<td></td>
</tr>
<tr>
<td><strong>PSYCHONEUROSES:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46*</td>
</tr>
<tr>
<td>Anxiety</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td>42*</td>
</tr>
<tr>
<td>Hysteria</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situational</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>42*</td>
</tr>
<tr>
<td>Unclassified</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total psychoneuroses</td>
<td>24</td>
<td>46*</td>
<td></td>
<td>46*</td>
<td>49*</td>
</tr>
<tr>
<td><strong>PERSONALITY AND CHAR. DISORDERS:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotionally instable</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inadequate personality</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td>10*</td>
</tr>
<tr>
<td>Schizoid personality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46*</td>
</tr>
<tr>
<td>Pers. pattern disturb.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pers. trait disturb.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociopath</td>
<td>38*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total pers. and char. disorders</td>
<td>74</td>
<td>46*</td>
<td></td>
<td>46*</td>
<td>58*</td>
</tr>
</tbody>
</table>

**Legend to Table II**

- Type of medical facilities:
  1 = Univ. hospital psychiatric emergency dept.
  2 = Observation unit (inpatient)
  3 = University hospital psychiatric clinic
  4 = Psychiatric hospital
  5 = Private practice
  6 = Psychiatric ward, general hospital
  7 = Follow-up in community
  8 = Outpatient dept. of univ. hospital

- Type of patients:
  A = Naval psychiatric patients
  B = Inpatients
  C = Clinic outpatients
  D = Private patients

- Type of classification:
  X = APA
  Y = Special Navy classification
  V = Langfeld's adaptation of Kraepelin
  W = Amended ISC

- Interval between contacts:
  1 = 1 week to 4 weeks
  2 = 2 weeks to 4 weeks
  3 = 1 day to 26 weeks
  4 = 7 to 10 years

- Time of diagnosis:
  ad. = Admission
dl. = Discharge

- Symbols:
  * = Includes organic psychoses not otherwise specified
  † = Combination of alcoholism, acute brain syndrome, and alcoholic intoxication
  ‡ = Excludes psychotic depression
  § = Includes symptomatic and confusional psychoses and those with no diagnosis specified
  Δ = Includes involutional psychosis
  # = Reworked data

- Not specified
An examination of Table II indicates that the stability of diagnosis for specific categories displays a rather wide degree of variation. The individual categories for the Organic Psychoses are on the whole fairly consistent. For chronic brain syndrome, 92 per cent; 72 per cent and 92 per cent for epileptic psychoses; 89 per cent for general paralysis; 70 per cent for mental deficiency; 64 per cent for the alcoholic psychoses. On the other hand, the remaining categories have a low consistency: 53 per cent for senile dementia and 29 per cent for cerebrovascular psychoses.

Consistency in the Functional Psychoses is of a low order. The highest is 87 per cent for Langfeldt’s “nuclear schizophrenia,” with 69 per cent for manic depression (all forms) being next. Langfeldt also reports a consistency of 55 per cent for his schizophreniform psychoses; for the schizoaffective category it is 28 per cent. Consistencies for the affective psychoses are 43 per cent and 46 per cent; mixed psychoses (unspecified) 31 per cent; and paranoid psychoses 29 per cent. The lowest reported consistency is 10 per cent for the involutional psychoses.

Reported consistencies for subcategories of Psychoneurosis are very low, ranging from 0.0 per cent for situational to 42 per cent for neurotic depression.

Consistency for Personality and Characterological categories tends to be uniformly low. The one exception to this is the reported 62 per cent for schizoid personality. The range of consistency for the other categories is from 10 per cent for personality pattern disturbance to 47 per cent for emotional instability.

A further examination of this table in terms of general diagnostic categories reveals that consistency for them is also rather low: for the “total” organic psychoses, 53 per cent and 65 per cent; for “total” (unspecified) schizophrenia, from 37 per cent to 70 per cent; for the “total” functional psychoses, 43 per cent and 59 per cent; for “total” psychoneurosis from 24 per cent to 49 per cent; for “total” personality and characterological disorders, from 46 per cent to 74 per cent.

In summary, a review of Table II indicates that for the specific categories, with the exception of several of the organic, the consistency over time of diagnosis is for the most part low. Also, the broad diagnostic categories appear to display a low order of consistency over time, with the diagnoses of schizophrenia, and those of characterological disorder, showing the greatest consistency.

Frequency in diagnostic categories for comparable samples.—The agreement of the distributions of diagnoses for two samples drawn from the same population is another indication of the reliability of the diagnoses. If the two distributions differ no more than one would expect from sampling errors, it may be concluded that the diagnostic process which underlies the classification system is reliable. Even if they differ significantly by statistical tests, but do not differ widely numerically, or, if categories maintain the same rank order in the two samples, some evidence of reliability may be inferred. The given comparisons may be made between comparable
### TABLE III

**Frequency (per cent) in Diagnostic Categories for Comparable Samples**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Med. facility*</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sex</td>
<td>F</td>
<td>MF</td>
<td>MF</td>
<td></td>
</tr>
<tr>
<td>Class. type b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison*</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>Significant diffs.</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Group N</td>
<td>141 206 191</td>
<td>75 68 48</td>
<td>128 166</td>
<td>554 606</td>
</tr>
</tbody>
</table>

| ORGANIC PSYCHOSES:                |                       |                 |                    |                 |
| All organic                       | 9                     | 7               | 10                 | 11 10 4         |
| Acute brain synd.                 | 13*                   | 15*             | 6*                 | 9*              |
| Chronic brain synd.               | 7*                    | 13*             | 24*                | 27*             |
| Mental deficiency                 | 1*                    | 2*              | 3*                 | 5*              |

| FUNCTIONAL PSYCHOSES:             |                       |                 |                    |                 |
| Affective disorders               | 6                     | 5               | 4                  | 2*              |
| Manic depressive                  |                        |                 |                    | 1*              |
| Depressions, including psychotic & | 15*                   | 14*             |                    |                 |
| & neurotic                        |                       |                 |                    |                 |
| Involutional                      | 36                    | 23              | 26                 | 29 22 67        |
| Schizophrenias                    | 10*                   | 8*              | 16*                | 22*             |
| Paranoid state                    | 1*                    | 0*              |                    |                 |

| PSYCHONEUROSES AND CHAR. DISORDERS: |                      |                 |                    |                 |
| Psychoneurosis                    | 30                    | 45              | 30                 | 35 40 8         |
| Psychoneurosis, except depression | 13*                   | 16*             |                    |                 |
| Anxiety                           | 6*                    | 6*              |                    |                 |
| Conversion                        | 5*                    | 6*              |                    |                 |
| Obsessive-compulsive              | 1*                    | 3*              |                    |                 |
| Other neurosis                    | 1*                    | 1*              |                    |                 |
| Psychophysiological disorder      | 4*                    | 1*              |                    |                 |
| Situational reaction              | 2*                    | 2*              |                    | 5*              |
| Char. pers. disorder              | 12                    | 13              | 22                 | 12 16 16        |
| Passive-aggressive                | 25*                   | 19*             | 14*                | 11*             |
| Sociopath                         | 3*                    | 3*              |                    |                 |
| Other pers. disorder              | 7*                    | 4*              |                    |                 |
| Other                             | 7                     | 7               | 9                  | 5 7 10          |
| No diagnosis                      | 9*                    | 10*             |                    |                 |

---

**Legend to Table III**

- **Symbols:**
  - `*` = Reworked data
  - `≈` = Not specified

- **Type of med. facility:**
  - 1 = Psychiatric hospital
  - 2 = Psychiatric consultant service, general hospital

- **Type of classification:**
  - 1 = Abbreviated APA

- **Kind of group comparison:**
  - A = Among wards in one hospital
  - B = Among administrators on one ward
  - C = Between years in one service
  - D = Between two independent psychiatric units taking patients on alternate days from same population
  - E = Between psychiatrists, patients assigned in turn

---

Wards in one hospital, between samples diagnosed by different psychiatrists, or between the proportion of assigned diagnoses from year to year. The primary statistical assumption underlying such studies is that there has been random selection of the patients in the samples under comparison. The data are shown in Table III.

An analysis of Table III reveals significant differences in the frequency with which some diagnostic categories were utilized.
In the Pasamanick study (49), female admissions were randomly distributed across three comparable wards. They were found to be homogeneous with regard to age, residence, education, type of admission, and marital status. Utilizing the base rate for each diagnosis determined from the combined samples, a statistically significant value of $\chi^2$ was obtained for the total contingency table, indicating an effect of ward on diagnosis. When the distribution of individual diagnostic categories was examined, discrepancies across samples were found to be marked for schizophrenia, psychoneurosis, and characterological disorders. In one ward that had had three different administrators, a marked discrepancy was found between the distribution of diagnoses by one administrator and those of the other two. The chief discrepancies were for schizophrenia and psychoneurosis.  

Kreitman (31), in his review of Pasamanick's study, after excluding the comparisons with the small sample of 48, which he regarded as too small, concludes that the only categories which show wide disagreement are psychoneurosis and character disorders. For the remaining categories no statistically significant differences occur.

In another study (47) which compares two independent psychiatric units, significant differences were found in diagnoses of acute brain syndrome, involutional psychoses, and schizophrenia. And finally, in still another study (41), although categories of diagnoses were combined rather broadly, significant differences were found (for both male and female services) between the distributions of diagnoses made by different psychiatrists to whom patients were randomly assigned.

Of the four studies summarized in Table III, only that of Wilson & Meyer (73) showed good agreement in distributions.

In general, the results of comparative studies of random samples with regard to distribution of diagnoses do not yield a consistent picture regarding reliability.

Validity

Validity deals essentially with the question of how well diagnosis attains its goals. We have already pointed out that no overall answer regarding validity can be given because the goals of diagnosis are so many and so disparate. That is why the validity of a given diagnosis is one of the most baffling problems that we face. We have already pointed out that the three major goals of diagnosis are actiology, prognosis, and selection of therapy.

In the discussion of validity in psychometrics the following types of

$^9$ Although the value of $\chi^2$ for the total contingency table was statistically significant, a more detailed analysis by categories indicated statistical significance for only two of them: schizophrenia and psychoneurosis. The statistical method used in performing the more detailed analysis on these contingency tables is a generalization of the Scheffé method of multiple comparisons [Scheffé, H. The Analysis of Variance (New York, Wiley, 1959) and Marascuilo, L. A. Large sample multiple comparisons. Psychol. Bull., 65, 280-90 (1966)]. The author is grateful to Joseph Fleiss for suggesting this type of analysis.
validities have been proposed: concurrent validity, predictive validity, construct validity, and content validity [(1); (14, pp. 96-125)]. Validity is defined as the degree to which a test measures what it purports to measure, and usually the criterion for such validity is singular. Whether we can utilize this type of approach to assess validity in the field of diagnosis is debatable, but it is worth a trial and this section will deal with this attempt. Predictive validity has, of course, its counterpart in prognosis and follow-up to see whether the prognosis was borne out. Concurrent validity has its counterpart in the degree to which methods other than the clinical interview corroborate the clinical diagnosis. Here, of course, the results of such techniques as the MMPI, Rorschach, and similar techniques can be matched against the clinical diagnosis. Construct validity has its counterpart in the attempt to see whether such aspects of psychopathology as can be measured by objective approaches (thought disorder, flatness of affect, or anxiety) can be related to a diagnosis which depends upon the presence of such psychopathology. Content validity can be approached by noting whether the characteristics of a given disorder are actually inquired into when a given diagnosis is made. The preparation of systematic and structured interviews provides an opportunity for examining content validity.

Concurrent validity.—One of the chief methods utilized in confirming diagnoses through other means than the clinical interview is the use of self-reporting techniques, such as the Minnesota Multiphasic Personality Inventory (MMPI) or the Maudsley Personality Inventory and projective techniques. The self-reporting techniques have demonstrated their usefulness for screening purposes during World War II and in counselling centers. In fact, they seemed to be somewhat more consistent measures in military screening when the comparison was made between the average scores attained in the various military camps and the rate of psychiatric rejection in these camps (61). From camp to camp the variation in scores was much smaller than the variation in rejection rate. However, for individual diagnoses of mental patients in hospitals and clinics, these instruments have not always proved themselves so useful. The degree to which specific individual diagnoses are corroborated by the MMPI and similar techniques is unknown since, for the most part, uncontaminated comparisons are rarely found.

The history of Rorschach validity studies also leaves much to be desired. In a careful analysis of all the available data, Zubin, Eron & Schumer (85) concluded that where studies were conducted in which contamination was eliminated, the correspondence between Rorschach based diagnoses and clinically based diagnoses was not very high. Furthermore, after weighing the evidence for the usefulness of the Rorschach technique in diagnosis, they concluded that only when the protocols are regarded as interview material can any relationship be established between the Rorschach results and those of the clinical interview. For this reason one wonders why the ink blots are necessary—why not interview directly without their
aid and introduce systematic structured interviewing as a substitute. It should be remembered that the self-reporting devices are also derivatives of the interview but they do not allow for systematic probing of the yes-no answers possible in interviewing. The primary advantage of the interview over its derivatives inheres in the fact that the chief purpose of interviewing in psychopathology is not to obtain factual information but to determine attitudes and feelings (79, 86). The only means that the interviewer has available for eliciting feelings and attitudes is the utilization of the tactics and strategies of verbal behavior, e.g., reinforcement, surprise, change of tempo, silence, etc. Although the scientific investigation of these tactics and techniques has barely begun, it is quite clear that their availability to the interviewer gives him an advantage over self-reporting inventories and that skill in these techniques separates the good from the poor interviewer.

Concurrent validity of diagnoses emanating from biochemical, physiological, sensory, perceptual, and psychomotor techniques are still so embryonic that they need hardly be mentioned except in so far as they contribute to construct validity; they will be discussed later.

Thus, concurrent validity, which attempts to answer the question how well do other approaches agree with the results of the clinical interview, is not too well-established in the field of diagnosis.

**Predictive validity (prognosis).**—There are three aspects of predictive validity in the area of diagnosis: (a) prediction of outcome of illness for a given diagnostic group; (b) selection of therapy to maximize good outcome; (c) prediction of other specific characteristics and behaviors from the diagnosis. It should be noted at the start that there are very few if any truly prognostic studies. Most of the studies to be reported here are epiphenomenal, in that they are postdictive rather than predictive.

The most striking example of predictive validity in the field of mental disorders is Kraepelin's attempt to include outcome as part of the diagnosis in dementia praecox. If he had succeeded, we would have had a paradigm of predictive validity. More successful is the predictive validity attached to manic depressive psychosis. Here the past history of the patient is one of the most basic features in making the diagnosis, and the reoccurrence of episodes is one of the most striking instances of predictive validity. Zubin et al. (80) collected 364 studies available in the literature as of 1956, in which diagnosis was related to outcome of the subcategories of schizophrenia. Each study was tabulated with regard to its finding on the relationship between diagnosis and outcome (outcome either favorable, unfavorable, or nondefinitive). The following diagnoses were found to have a good prognosis: catatonic schizophrenia with 119 studies indicating good outcome and two studies showing poor outcome (119/2); mixed schizophrenia (5/1); atypical and undifferentiated (15/0); reactive schizophrenia (2/0); schizo-affective (3/0); and schizo-depressive (1/0).

The following diagnoses were found to have an unfavorable outcome: simple schizophrenia (1/33); paranoid schizophrenia (55/39); catatonia in
later life (0/15); hebephrenia (13/57); and process schizophrenia (0/3). It should be noted that with only a few exceptions, the introduction of the somatic therapies made but little alteration in these predictions.

Diagnosis of patients in London mental hospitals was found to be highly associated with outcome in the classic study of Norris (48). High rates of release were found for manic depressive psychoses and the neuroses, and rather low release rates for schizophrenia and geriatric patients.

For schizophrenia only about one third were not likely to recover sufficiently ever to leave the hospital within five years from their admission. These figures are not too different from those reported in the U.S.A. Of the manic depressives, only 10 per cent remained continuously hospitalized whereas of those discharged approximately 40 per cent have one readmission and about 8 per cent a second readmission.

For the neuroses, 95 per cent were released and around 5 per cent remained in the hospital longer than two years. This fits well with the findings of Kessel & Shepherd (28) and Shepherd & Gruenberg (57), in which a comparison of the age-specific incidence and prevalence rates in several countries indicated that the average duration of neurosis was about two years.

Additional evidence for the rather brief average duration of neurosis is afforded by a recent study of Greer and Cowley (18a) reporting the findings on 175 inpatients discharged from the Professorial Service at the Maudsley Hospital in London during the 3-year period 1953-55 and whose admission diagnosis was neurosis. Presumably all psychoneurotics eventually are discharged so that the data for the discharged patients can be taken as typical. The median duration of hospitalization was 12.5 weeks (12.2 weeks for those showing any improvement and 15.0 weeks for those showing no change or worsening). The median duration of symptoms prior to hospitalization for the key admission was 12.7 months. Hence, the total duration of illness can be estimated at about 16 months. About 41.1 per cent had had their symptoms for more than 2 years prior to admission and some of these for their entire life, but since even of these hard-core neurotics only 17.8 per cent failed to show any improvement, it may be concluded that for the vast majority of treated neurotics, the average duration of total illness is from 1.5 to 2 years. How this duration would be changed if we considered untreated as well as treated neurotics remains to be investigated. However, whether the group of patients treated at the Maudsley is a good representation of treated neurotics in general is an open question.

Of Norris' geriatric group, though death was the chief avenue of release, the proportion of patients in their seventh decade of life who are released from the hospital varies between 20 and 50 per cent, the patients with affective symptoms showing the brighter prospects. During the eighth decade of life the prospects are grimmer, the proportion of releases varying from 6 to 30 per cent.

Since most of the follow-up studies are based on the patient's presence
in, or absence from, the hospital on the day of the follow-up, attempts have been made to develop indices that would incorporate more of the data than those available as of the day of follow-up. Crandell et al. (12, 13) developed measures which they successively named the “occlusive” index and the “mobility” index; most recently Burdock & Hardesty (8) further developed this idea into the “outcome” index, in which the proportion of days out of the hospital during the follow-up period is multiplied by a function of the number of releases and number of readmissions. This index, which varies from 0 to 1, divides the group into three rather distinct subgroups—those who never get out, those who come in for a short period, leave, and never return, and those who oscillate in and out of the hospital.

Also, since the outcome index may not reveal the degree of readjustment the patient has undergone, it is interesting to know how well the released patient is doing in the community. Such an investigation was made by the Veterans Administration (18b). In a sample of male veterans consisting of 95 per cent schizophrenics under 60 years of age at admission, without any organic or somatic involvement, only one third were found to be working full time and half were not working at all.

In contrast, Wing (74) found about two thirds of the released patients working or managing households. However, his study tends to agree with the finding of Peterson, and also of Norris, that from one fourth to one third of the first admissions for schizophrenia remain chronically ill.

One of the most important aspects of diagnosis is selection of therapy and, here, the choice is made usually on the prediction of outcome, which can be taken as an example of predictive validity. There are any number of studies relating diagnosis to outcome of treatment; Zubin et al. (80) found that the generally dire outcome of hebephrenic schizophrenia was altered by the shock therapies. Whereas before the development of the somatotherapies the ratio of studies showing good prognosis was only 1/30, under the somatotherapies, excluding psychosurgery, the ratio changed to 12/19. Similarly, the ratio for paranoid schizophrenia changed from 8/31 to 47/8. Caution must be used in interpreting these data since there is no guarantee that the diagnostic bases remained unchanged from the preshock to the shock period.

Kreitman et al. (32) examined the relationship between diagnosis and recommended therapy. They found that the degree of agreement in choice of therapy between two psychiatrists varied with diagnostic category. For the functional psychoses, the degree of agreement on choice of therapy was 77 per cent, for neurosis 54 per cent, for organic disease 51 per cent, and for other diagnoses 33 per cent. The degree of agreement on diagnosis was generally higher than on treatment.

Bannister, Salmon & Leiberman (4) investigated the relation between diagnosis and treatment over the entire range of neurotic, psychotic, and organic patients. Utilizing a contingency table (diagnostic categories versus 14 specified therapies), they ascertained whether treatment had been pre-
scribed in a nonchance manner. They found, as one might expect, that more neurotics received psychotherapy, more psychotics electroconvulsive therapy (ECT), and more organics phenothiazines or nonspecific treatments. When a specific breakdown of the diagnoses was made, the trends found in the major groupings were confirmed but some additional facts emerged. More schizophrenics were given phenothiazines and "long" ECT treatments, more depressives (affectives) antidepressants. More reactive depressives and more of the affective group were administrated antidepressants; though a few of the former also received psychotherapy, none of the latter group did. Hystericls differed from anxiety states with regard to nonspecific treatment, the former receiving more than was expected from their base rate. There was a tendency for personality disorders to receive psychotherapy. Addictions (mainly alcohol) were primarily treated with phenothiazines with or without psychotherapy. When age and sex influences are eliminated statistically, the picture does not change in any important respect.

The authors point out in summary that though there are statistically significant associations between diagnosis and type of therapy, the number of "correct" choices of therapy (correct meaning the most popular therapy for a given diagnosis) occur 182 times out of 1000 (as compared to 71 based on chance) when 14 choices of therapy are presented and three major diagnostic categories (neurotic, psychotic, and organic) are utilized. This, though statistically significant, still leaves 818 instances of "incorrect" choices. If we use more specific diagnoses the number of "correct" choices goes up to 329. If the treatment categories are grouped into ECT, chemotherapy, psychotherapy, nonspecific and other, the number of "correct" choices of therapy goes up to 500 as compared to a chance expectancy of 200. The authors conclude that "... although diagnostic decision and choice of treatment in psychiatry are significantly related for some categories, the wide areas of 'no association' suggest that habitual or logical links are very limited in extent. ..." Whether the advent of drug treatments produced better outcomes in schizophrenia is still a moot point. Peterson & Olson (51) compared outcome in chronic schizophrenia before the advent of the drug era with the results attained in one hospital during the drug era. Since they had no predrug data on first admissions in this hospital they utilized the well-known data at the Warren State Hospital for the predrug era (1936-45). They concluded that on follow-ups of five years or longer, about 25 per cent of the first admissions are still in the hospital either continuously or through readmission. This is not different from the data for the predrug era and hence, though drugs may produce more oscillation in the patient population, hospital stay of long-range, hard-core schizophrenics is unaffected by the drugs.

Construct validity.—The specific constructs which underlie the diagnostic schema, and which could become the basis for testing the construct validity of a diagnostic grouping through observation and experimentation,
constitute an area which is too vast to permit even enumeration, much less evaluation. For this reason, the area of construct validity should be a chapter in itself in some future Annual Review. Here we can only point to a few outstanding examples. Flatness of affect is a characteristic often used in the diagnoses of schizophrenia. Since the rating of affect is very difficult, Salzinger, Portnoy & Feldman (52) undertook to devise a specific technique for eliciting and measuring the frequency with which affective utterances can be elicited from schizophrenic patients and normals. The technique consists essentially of using reinforcement to encourage the emission of affective utterances in a sequence of operant level, conditioning, and extinction periods. The technique is highly reliable, indicating that schizophrenics, if they can be engaged in conversation at all, do not differ from normals in operant level nor in conditioning but they do extinguish more rapidly than normals. Furthermore, schizophrenics who show higher rates of conditioning tend to get out of the hospital sooner.

Communicability of schizophrenic speech, which is another aspect of schizophrenic behavior, has been successfully measured by means of the cloze technique (52). Thought disorder, especially that aspect which is known as overinclusion, has been studied by Payne (50), who postulates that the essential difficulty is "an impairment of some central filtering process whose function it is actively to inhibit external sensations and internal thoughts which are not relevant to whatever is the focus of attention at a given time." This disorder is found in only 50 per cent of schizophrenics but can be induced in normals by LSD and mescaline. It is a passing phenomenon disappearing early in the course of the illness and therefore does not tend to occur in chronic. Garmezy (18) has provided measures of two facets of schizophrenic reactivity: "(1) inflexible and persistent withdrawal from psychologically noxious events and (2) a lessened responsiveness to social rewards." Garmezy tends to attribute these characteristics to early experience with an ambivalent mother and thus connects up his findings with the "double-bind" of Bateson et al. (5) and the familial interaction of Lidz and associates (37, 38).

The processing of stimulus input by schizophrenics and other behavior disorders has attracted the attention of workers like Shagass (56), Callaway (11), Sutton & Zubin (63), Venables (67-70), and many others. It is quite clear that the physiological and sensory responses of the mentally ill provide an important possibility for construct validity investigations. These will be touched on in the final section.

Emerging Trends

That there is widespread dissatisfaction with current diagnostic procedures is very clear. This dissatisfaction has presented a challenge to psychopathology because now, if ever, good diagnostic procedures are essential. When all we had to offer the behaviorally disturbed person was custodial care, there was not much need for diagnosis, but, paradoxically, the
heyday of diagnosis occurred during this custodial period. Diagnosing was perhaps more an academic exercise in search of etiology than a practical tool. Today, with the bulging armamentarium of therapies, we must find some way of selecting the best therapy for each patient. But just now, diagnosis is at its lowest ebb. Furthermore, medicare and the opening up of community mental health centers will bring to our facilities many self-selected false positives whose diagnoses will be most essential for safeguarding not only the patients from unnecessary or ill-advised treatment but also the facility from being needlessly overrun.

In response to these needs, three points of view are emerging. The first, a somewhat defensive view, states that the current diagnostic system is excellent and can meet the new challenges and that the only needed improvement is in better training of the diagnostician. At the opposite end of the spectrum is the view that behavior disorders are nothing more than the response of an essentially normal individual to severe stresses and strains and that no diagnosis as such is possible or even desirable. Somewhere in between is the view that the current diagnostic system represents a good starting point from which to improve approaches to classification.

Several studies have examined the sources of disagreement in the current diagnostic approach. Ward et al. (71) found in an experimental investigation nine basic sources of disagreement in diagnosis. They attributed these to (a) the inconstant behavior of the patient (5 per cent), (b) inconstant behavior of the diagnostician (32.5 per cent), and (c) inadequacy of the nosological system (62.5 per cent). Apparently both the clinician and the diagnostic schema bear the brunt of the blows. If we could introduce more systematic structural approaches to interviewing, so that each clinician would cover the same ground with each patient as far as is consistent with clinical needs, we would stand to gain almost 33 per cent increase in accuracy or, at least, in agreement. The shortcomings of the nosological system can, of course, be reduced by carefully overhauling the diagnostic schema. With the reduction of these two major sources of error, agreement, at least, could be enhanced.

But what about validity? Some idea of concurrent validity can be obtained by examining the syndrome or syndromes which presumably characterize each diagnostic group. Freudenberg & Robertson (17) found considerable overlap of symptoms among the various diagnostic groups. Zigler & Phillips (77) also found such low relationships between symptoms and diagnoses that they concluded: “membership in a particular diagnostic group conveys only minimal information about the symptomatology of the patient.” Similarly, Katz, et al. (27b) point out that diagnostic categories are not specific for configurations or “patterns of symptomatology.”

The absence of relationships between diagnoses and symptomatology in a system where symptomatology forms the backbone of the diagnosis is a puzzling phenomenon indeed. It is possible that the diagnosticians observe different things in different patients and, despite these differences, still give
these different patients the same label. What is it then that they base their judgments on? Is it some intuitive process which defies description? To consider this, we must provide reliable methods for patient interviewing, methods for estimating symptoms, and methods of scaling them into significant dimensions; then we must provide individual profiles across these dimensions.

The attempt to provide a dimensional approach to diagnosis dates back to Father Moore (45), who was the first to apply a factor analysis to patient symptomatology, thus supplying confirmatory evidence for the Kraepelinian schema, with some minor emendations. Since then, quite a number of research workers have applied themselves to the problem of dimensionalizing the behavior of patients suffering from behavior disorders. These include Degani, Jenkins, Wittenborn, Eysenck, Cattell, Geurtin, Lorr, Cohen, Zubin, and Overall. A review of this area has appeared in Lorr, Klett & McNair (40). Here we can give only a bird's eye view of the approach, as exemplified in the work of Lorr, one of the more assiduous researchers in this field.8

Lorr starts with the current clinical approach in which the natural features of psychopathology are revealed through clinical interviewing and expressed in the concepts and language of clinical psychiatry. He thus utilizes the clinician where the clinician's chief strength lies—as an observer—and in ways natural to him. These observations are expressed numerically in scaled judgments on a series of items characterizing the patient's behavior. By factor analytic means, these numerous item scales are reduced to the larger, more general dimensions implicit in them. After many studies of various items and samples, he has settled on some 75 item scales, which he has reduced to ten first-order factors, called syndromes, and three second-order factors. These are then treated in conventional psychometric fashion—transformation to common standard score and centile units of the syndromes, and norms. A ten-element profile results. This captures most of the non-error descriptive variance in symptom manifestation and offers it in a form amenable to quantitative analysis.

By objective multivariate procedure, samples of profiles are analyzed to yield mutually exclusive subgroups of similar profiles, or types. Lorr has identified about a dozen such types [see his Explorations in Typology—now in press—but six were described in Syndromes of Psychosis (40)]. Each type can be delineated in terms (e.g., means or distributions) of the ten "syndrome" elements.

Further study within the schema is possible. For example, he subjected the data to a multiple discriminant function analysis to determine the number and nature of the composite dimensions (canonical variates) which significantly differentiate among the types. This procedure can be used practically to classify patients into types in a way which minimizes error,

8This summary of Lorr's work is based essentially on an overview by Dr. Jacob Cohen, to whom the author is grateful.
i.e., to assign diagnoses, provided the types can be used to define diagnostic categories.

Much of this is still provisional, but it may be considered as an objective translation of what the great descriptive psychiatrists Kraepelin, Bleuler, and Meyer sought to do. One can't help but contrast a research program which seeks a diagnostic schema such as Lorr's with the majority vote procedure by which the American Psychiatric Association arrived at its current nosological schema!

One thing more must be said. It is conceivable that after such a schema is in final form, and one has done the necessary work to determine its antecedent (aetiologic), concurrent, and consequent (prognostic) correlates, it may prove a bust! There is, of course, no guarantee that the symptoms available for observation in a conventional interview are not epiphenomena, only distantly related to the essence of psychopathology, i.e., etiology, course, outcome. We may yet be displaced by biochemists. But I am not prepared to proceed on such an assumption. The present state of diagnosis provides no evidence. Until the symptom picture has been properly (objectively, scientifically, mathematically) exploited through such programs as Lorr's we will not know.

A more recent development in the dimensionalization of diagnoses is the introduction of systematic structured interviews (9, 10, 59, 88). With a common set of questions and items provided for the interviewer, the results can be codified into a total score representing the severity of illness and subscores representing the status of the patient on a variety of dimensions. The usefulness of the structured interview has been demonstrated in an experiment, conducted by Katz and his colleagues (87), in which a group of diagnosticians were asked to indicate the presence or absence of psychopathology on behavioral items during the presentation of a videotape of an interview. They were also asked to make a diagnosis. Of the group of 35 participants, 21 labeled the patient psychotic and 14 labeled him neurotic. An examination of the average profile given the patient by those who diagnosed neurosis as opposed to those who diagnosed psychosis indicated that the profiles were quite alike except with regard to one dimension, apathy. Those who perceived much apathy called the patient psychotic whereas those who did not called him neurotic!

The major problem now before us is—of what use is the diagnosis, even when obtained by means of systematic structured interviews, if it does not correspond to the symptom patterns provided by the dimensional approach? Furthermore, of what use is diagnosis if individuals with the same diagnostic label do not respond similarly to a given therapy or do not attain similar outcomes on follow-up?

One solution to this problem is the biometric approach (87). There are basically two other approaches, the clinical and the phenomenological. The clinical approach to diagnosis, with its focus on treatment and its search for etiology and prognosis, has already been discussed. The phenomenological
approach, which one finds in such masters as Jaspers, has also been touched upon in so far as it is concerned with the characteristics of the patient as perceived by a well-trained naturalistic observer, i.e., a description of his total behavior including subjective feelings and inner experience as perceived by the inner world of the phenomenologist and reported in descriptive terms. The third approach is the biometric, which tries to make the phenomenological description public in such a way that it becomes reproducible in the hands of other observers. The clinical approach depends primarily on the concept that specific entities of mental disorder exist and that the patients belonging to a given category have certain characteristics in common which are not shared by other groups. Thus, a certain discontinuity is postulated in the characteristics of the patients belonging to a given category. This discontinuity may refer to one characteristic, i.e., presence versus absence of hallucinations, or it may refer to a discontinuity in the relationship between several characteristics, i.e., a combination or pattern of a certain degree of thinking disorder, a certain level of affect, and a certain degree of interpersonal contact as a syndrome of schizophrenia. It is interesting to inquire whether the discontinuity inhere in the behaviors of the patients or in the behavior (judgment) of the observer. If it inheres in the patients, there is some hope of finding statistical methods for establishing the presence of nonoverlapping categories of disorder. However, if patient behavior is continuous with regard to the monovariate and multivariate distribution of their characteristics, we are faced with the problem of determining the threshold for discontinuity in the judgment of the observer. In the latter case our diagnostic procedures become highly dependent on the sociocultural milieu of the patient and schooling of the diagnostician. Thus, while the clinical approach assumes the presence of discontinuities, the phenomenological and the biometric approach do not necessarily make this assumption and can operate with continuous (dimensionalized) phenomena or variates as easily as with discontinuous ones. It is interesting to speculate whether the contrast between the clinical typological and the dimensional approach are fundamentally different in essence. It is amusing to realize that the apple knocker who was utilized to separate the good apples from the bad apples has now been replaced by an automated process which depends upon the relative absorption of pairs of monochromatic light waves. Thus, a judgmental discontinuity has been replaced by a dimensionalized procedure dependent upon continuous dial readings. The conflict between typology and dimensionality may be a pseudoconflict dependent upon the state of knowledge of the field. If we now find some genetic underpinnings for good and bad apples, we might revert to typology. Later, however, when the enzymatic qualities of the alleles of the genes are examined and their rate of chemical activity determined, dimensionality may be reintroduced and thus the pendulum will continue its swings (87).
The biometric approach begins with the assumption that the diagnosis and classification of behavior disorders will become scientifically acceptable only after objective measurement of patient behavior is introduced. In order to introduce such quantification we must find a suitable scientific model as a framework on the basis of which the measurement can proceed. The second assumption is that the ultimate goal of diagnosis is etiology since without it our measures are only tentative and our treatments only temporizing. What scientific models are available for the etiology of behavioral disorders? The following have been described (82, 84): the social-cultural model, the developmental model, the learning theory model, the genetic model, the internal environment model, and the neurophysiological or brain function model. To evaluate the importance of each of these models in the diagnosis of a given case, measurement techniques suitable to each of these models are required. The models have been divided into three groups: (a) the social-cultural, (b) the developmental and learning, and (c) the genetic, internal environment, and brain function.

For the first, the social-cultural model, culture-dependent techniques in the form of the systematic structured interview and objective observational techniques are provided. For the developmental and learning theory model, culture-fair techniques are provided and for the genetic, internal environment and brain function models, culture-free techniques are being developed.

The culture-dependent techniques are aimed primarily at measuring deviation from expected social-cultural norms, which constitute the most important bases for the detection of behavior disorders. It is clear that such measures will show variation from culture to culture and even between subcultures in the same region. The systematic interview interpreted in the light of local expectancies is the tool most suitable for this area.

The culture-free techniques are aimed essentially at discovering deviations in behavior which would characterize an individual regardless of his cultural milieu. An example would be the biochemical techniques for detecting phenylketonuria (PKU). Behaviorally, the first 1000 msec interval following stimulation under laboratory conditions has been found promising in the detection of deviations from normals in schizophrenics (63) and depressives (56). There is a growing literature in this general area (11, 29, 30, 67, 68, 69, 70, 76). A single example will suffice. Shagass (56) has reported separation between psychotic depressives and normals, with almost no overlap, using the recovery time of the evoked potential to the second of two somesthetetic stimuli. Interestingly, as the patients improved clinically the evoked potential recovery time approached that of the normals.

The culture-fair techniques are aimed at detecting behavioral deviations from certain transcultural expectancies, such as greeting behavior, communicability of speech, etc. Here, translations have to be made from culture to culture to obtain the equivalence of the different culturally determined norms for communicability, greeting, grieving, etc. We have already
discussed the work of Salzinger on communicability and affective verbal behavior, of Payne on sorting behavior, and of Garvey on effect of reward and punishment.

The multiplicity of data provided by the biometric approach has to be integrated in some manner if it is to become useful. If the current diagnostic groupings, despite their relatively low reliability and validity, are adopted as tentative criteria, biometric profiles can be obtained for the patients in each diagnostic category. A grouping of the similar profiles will yield subcategories which, if found useful, can help in the development of more homogeneous subgroups that are so necessary in clinical research (11a, 27a, 48b). Another approach is to apply techniques for fractionating populations into subgroups (78, 83). However, as Fleiss has pointed out (see 83), unless certain assumptions regarding the nature of the distribution of traits are made, there is an infinite number of ways to perform the fractionation.

Recently, following the lead of cardiology and other specialties, W. G. Smith (58) of Cornell proposed the utilization of Bayes' Theorem in computerizing diagnosis. After obtaining from a group of 14 clinicians the probability with which each one of a series of 41 well-known symptoms of psychopathology occurs in each of 38 specific mental disorders, he determined first that the degree of agreement on the stereotypes for each disorder was rather high (71 per cent for the specific symptoms and 86 per cent for the specific diagnoses). Smith then chose 30 patients to be seen by a clinician who evaluated the presence or absence, as well as the intensity, of the 41 symptoms. The clinician's diagnoses were compared with the probabilities given by the computer for these 30 cases when the conditional probability for each symptom was fed into the program. The results indicated that the disorder with the highest probability obtained from the computer agreed with the clinician's diagnoses in 87 per cent of the cases. If the two top probabilities are taken together, the agreement reaches 97 per cent. A similar procedure was developed by Overall (48a).

The primary difficulties which this approach faces as described by the author are the assumptions that underlie the method: (a) independence of symptoms, (b) independence of disorders, and (c) equality of probability of occurrence of each of the 38 disorders. The method does demonstrate that under specified conditions a high level of agreement can be reached. The validity of the diagnoses and its usefulness remain to be demonstrated.4

4 It should be pointed out that the area of childhood disorders has not been touched on in this review. The explanation is twofold: first, the state of diagnostic procedures is even more incoherent in this field than in the adult field; and second, a fair treatment of this area would require the attention of a specialist in the area. This reviewer would like, however, to call the reader's attention to the following recent publications bearing on this area: Benton (6a); Jenkins & Cole (26); Birch (7); Fish et al. (16); Eisenberg (14a); and the Committee on Child Psychiatry of the Group for Advancement of Psychiatry (11b).
CLASSIFICATION OF THE BEHAVIOR DISORDERS

One major source of confusion in diagnostic work stems from the differences in points of view regarding the relationship between psychopathology and personality. The three possibilities in this relationship have been discussed elsewhere (89). Psychopathology may be regarded (a) as synonymous with the personality of the patient, (b) as an interference with personality, or (c) as independent of personality. At the present time, the evidence is equivocal with regard to the tenability of any one of these propositions, and differences in diagnoses often arise from difference with regard to this relationship. In view of this situation, the null hypothesis of no relationship may be adopted and double entry tables for the classification of patients along both dimensions may yield better classification. In a book which arrived too late to be included in this review (16a), this issue becomes focal in diagnosis.

The absence of relationship between diagnosis and symptoms referred to earlier [Freudenberg & Robertson (17) and Zigler & Phillips (77)] may be partially explained by the assumption that some of the symptoms are really basic personality traits and, hence, according to the null hypothesis, basically unrelated to diagnoses. The presence of such "symptoms," or rather traits, may influence the course of an illness, but have nothing to do with its existence or causation.
LITERATURE CITED


16. Fish, B., Wile, R., Shapiro, T., and Halpern, F. The prediction of schizophrenia in infancy: II. A ten-year follow-up report of predictions made at one month of age. In *Psychopathology of Schizophrenia*, 335-53 (See ref. 10)


18. Garmezy, N. The prediction of performance in schizophrenia. In *Psychopathology of Schizophrenia*, 129-81 (See ref. 10)


18b. Gurel, L. *Patterns of Mental Patient Post-Hospital Adjustment,


35. Leighton, A. H. My Name is Legion. Foundations for a Theory of Man in Relation to Culture (Basic Books, New York, 452 pp., 1959)


45. Moore, T. V. The essential psychoses and their fundamental syndromes. In Studies in Psychology and Psychiatry, III, No. 3 (Williams & Wilkins, Baltimore, Md., 128 pp., 1933)


50. Payne, R. W. The measurement and significance of overinclusive thinking and retardation in schizophrenic patients. In Psychopathology of Schizophrenia, 77-97 (See ref. 10)


52. Saltzinger, K., Portnow, S., and Feldman, R. S. Verbal behavior in schizophrenics and some comments toward a theory of schizophrenia. In Psychopathology of Schizophrenia, 98-128 (See ref. 10)


64. Szasz, T. S. The classification of


72. Wilson, E. B., and Deming, J. Statistical comparison of psychiatric diagnosis in some Massachusetts state hospitals during 1925 and 1926. *Schizophrenia, Statistical Studies from the Boston Psychopathic Hospital* (Reprint 2, 17-37, 1934?)


87. Zubin, J. *Clinical, Phenomenological, and Biometric Assessment of Psychopathology* (Presented at Symp.

88. Zubin, J., and Kietzman, M. L. A cross-cultural approach to classification in schizophrenia and other mental disorders. In Psychopathology of Schizophrenia, 482-514 (See ref. 10)