BIOMETRIC APPROACH TO
PSYCHOPATHOLOGY:

Abnormal and Clinical Psychology—
Statistical, Epidemiological, and
Diagnostic Approaches

Joseph Zubin, Kurt Salzinger, Joseph L. Fleiss,
Barry Gurland, Robert L. Spitzer, Jean Endicott, and
Samuel Sutton
Biometrics Research Unit, New York State Department of Mental Hygiene
New York, N.Y. 10032

FOREWORD

Joseph Zubin

This review consists not of the contribution of a single author, but of several of the
staff members of the Biometrics Research Unit (BRU) of the New York State
Department of Mental Hygiene. Since we believe that such fields as abnormal and
clinical psychology can no longer be reviewed adequately by one worker, this
review may mark a turning point in future reviews of similar fields. The BRU was
established in 1956 with the mandate to provide an objective assessment of the
behavior of mentally disordered individuals with the view of improving diagnosis,
prognosis, and evaluation of outcome of mental disorders, as well as promoting
investigations into etiology and prevention. This chapter reviews some of the
accomplishments of this endeavor on the part of the BRU staff, as well as by
similar staffs and research workers elsewhere.

This review has been limited to the following sections:
1. Behavioral Analysis—Kurt Salzinger
2. The Measurement and Classification of Depression—Joseph L. Fleiss
3. Psychiatric Epidemiology—Barry Gurland

621
4. Attempts to Improve Psychiatric Diagnosis—Robert L. Spitzer and Jean Endicott
5. A Psychophysiological Approach to Psychopathology—Samuel Sutton

My task is to introduce this survey and then to conclude it with an integrated summary.

Introduction

The field of abnormal and clinical psychology, which at one time was a relatively homogeneous entity, has separated into two distinct and only slightly overlapping disciplines, with the clinical wing far surpassing the abnormal in number of adherents and applications. Abnormal psychology can still be regarded as the basic experimentally oriented part of the endeavor, while clinical psychology can be viewed as the applied field largely invested in treatment and application (292). The apparent impotence of the standard clinical tests in prognosis (280, 295), their low reliability, their lack of value as independent instruments in diagnosis, and their dependence for validity on the criterion provided by the clinical interview has forced BRU into abandoning the use of routine clinical tests in research on diagnosis and prognosis and evaluation of treatment. Instead we turned our attention to the clinical interview itself—our ultimate criterion of psychopathology—and developed a series of systematic structured interviews which have proved to be reliable and valid (247, 252, 253, 255, 256). The interview techniques have served as a basis for making cross-national comparisons, as described in the section by Gurland; as a basis for computer diagnosis, as described in the section by Spitzer & Endicott; as a basis for developing a newer anatomy of psychopathology in the form of dimensions of psychopathology, as described in the section by Fleiss.

With one leg up on the problem of diagnosis, we then turned our attention to the problem of more direct examination of the behavior of the patient. Following the footsteps of general learning theory, Salzinger in his section applies the behavioral analysis approach pioneered by Kanfer and Saslow to the salient features of psychopathological behavior exhibited by the patient. Sutton, on the other hand, makes a more atomistic analysis of the laboratory-elicited behavior of the patient in his physiological, sensory, perceptual, psychomotor, and conceptual responses. To elicit these responses he applied the framework of the Mendelev-like table of Burdock, Sutton & Zbin (19), in which the following types of stimuli are applied: idling state (when no stimulus is applied); energy stimulus (when some relationship exists between intensity and response, e.g., greater latency for more intense stimuli); and signal stimuli (in which the response bears little relationship to the stimulus intensity but is related to the signal value of the stimulus based on either prior experience or wired-in propensities). In his section he develops the rationale for utilizing the paradigm of the Mendelev-like table offers certain research strategies to overcome the "ornerness" of mental patterns and demonstrates their use in differential diagnosis.

Up to this point we have dealt mostly with the description of abnormal behavior. Description, however, is not enough, for it never cured a disorder, nor has it of its own accord yielded us entirely lacking. To the form of scientific end, we must proceed in the crucible of and is still in its prelimin end of this chapter.

I. BEHAVIORAL

Kurt Salzinger

Spurred on by the fact that reliability and validity of their own kind of behavior therapies: paradigm and that bas this is not to say that is. On the contrary, the twofold focus of the therapy and, therefore, also the main business is identification of the stimulus. When used in original form, behaviorists are able to prepare operant conditioning stimuli (precede the undesirable primary S's) and conditions that directly involve the organism, not topographical conditions (the particu...
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY

623

of its own accord yielded any insight into the nature of the disorder. To go deeper we must appeal to etiology, even though our knowledge of etiology is skimpy, if not entirely lacking. To make up for our ignorance we must resort to "as if" causes in the form of scientific models. However, mere postulation of models is also not enough; we must proceed further to draw from them inferred hypotheses and test them in the crucible of observation and experimentation. This part of our endeavor is still in its preliminary stages, and we shall deal with it in the summary at the end of this chapter.

1. BEHAVIORAL ANALYSIS

Kurt Salzinger

Spurred on by the fact that psychiatric diagnosis continues to suffer the pains of low reliability and validity, behaviorally oriented therapists have started to codify their own kind of behavioral analysis. This analysis has been applied to both kinds of behavior therapies: that based on the classical or respondent conditioning paradigm and that based on the instrumental or operant conditioning paradigm. This is not to say that in practice one kind of conditioning occurs without the other. On the contrary, the two always co-occur (212). What differentiates the therapies is the focus of the therapist. It determines the kind of behavioral analysis engaged in, and therefore, also the kind of assessment procedure used.

The main business of the therapist following the respondent paradigm is the identification of the stimuli that act as conditional stimuli (CS) for the patient’s current state. When using desensitization, the therapist must, according to Wolpe’s (286) original formulation, also establish a hierarchy of aversiveness of stimuli in order to be able to present the CSs in an increasing order of anxiety productiveness.

The operant conditioners, on the other hand, has the task of identifying the discriminative stimuli (SD) in the reinforcement contingency (SD → R → SR) that precede the undesirable or desirable behavior (R), and those reinforcing stimuli (primary SRs and conditioned S’s) that follow them; in addition, the operant conditioner must estimate the strength of those behaviors examined within the reinforcement contingency.

Distinctive Features of Behavioral Analysis

Behavioral analysis focuses on those aspects of the individual and his environment that can directly be used in therapy. Respondent conditioning emphasizes the autonomic responses, while the operant paradigm emphasizes the behavior of the patient that directly engages the external environment. Behavioral analysis is functional, not topographical. More important than the presence of a fear are the conditions (the particular CSs) that elicit the fear (respondent conditioning).

1 The author is also affiliated with the Polytechnic Institute of New York.

The writing of this paper was supported in part by NIMH Grant Number 1 R01 MH 2490-01 and 02.
important than the individual's fighting behavior are the conditions under which it takes place, namely when, where, and to what effect (operant conditioning).

Behavioral analysis classifies behaviors rather than people. This is not to say, however, that behavior therapists restrict their work to the analysis of single behaviors. They take into account all relevant behaviors. For example, if such therapists discover that a particular individual is currently reinforced for undesirable behavior, they determine in addition what reasonable behavior the individual has in his repertoire; then, if it is thought advisable, that person can be retrained to emit the reasonable behavior to obtain the positive reinforcers earlier obtained by undesirable behavior.

Finally, while diagnosis is supposed to shed light on the cause of disorders (actually it rarely does so for psychiatric disorders), the relationship of behavioral analysis to the cause of the disorder is relatively unimportant. More significant than how the behavior started is what is maintaining it. In some cases this amounts to knowing the cause, but not usually. On the other hand, behavior analysis does assume that most types of behavioral abnormalities are produced through one of the learning paradigms. Even though some particular physical insult may be vitally involved in a particular abnormality, the interaction of that physical insult with the conditioning paradigms is held responsible for the current abnormal behavior (see 213 for this kind of explanation of the immediacy hypothesis and schizophrenic behavior).

The Behavior- Analytic View of Testing

Goldfried & Kent (76) and Mischel (159) explain that test responses are regarded as signs of behavior by the traditional approach and as samples of behavior by the behavior theory approach. Basically, behavior analysts prefer direct observation of the individuals in the actual situation, but they accept the creation of an experimental analog situation, including role playing and the use of self-report techniques. In the better assessment programs, the patient is taught how to observe his or her own behavior in an objective manner; unfortunately, more often than should be true, self-report is obtained through the clinical interview, which is beset by the artifacts of influence on the part of the interviewer (217–219).

The traditional approach to testing also carries along with it the assumption that the test responses reveal behavior patterns invariant with respect to the situations in which they occur. However, as long ago as 1928, Hartshorne & May (92) showed that the behavior of individuals depends in large part on the situation (the reinforcement contingency). Behavior is flexible in that it varies with environmental stimuli. Only the most extreme abnormality would produce the kind of invariance of behavior that traditional tests demand for their validity.

The use of self-report is often found to be a poor substitute for behavioral observation. One study (88, 216) focused on the relationship between communicability (behaviorally defined) and amount of contact in a group of people. Objectively observed amount of contact was the best predictor of communicability; self-reported contact was predictive only when considering presence vs absence; and self-estimated degree of closeness was not predictive to a sig-

ificant degree how much are specified, such specific, and cl.

Behavioral A

Although the its effect on measurement has spru behavioral as environment mensions as behavior set characteristic behavior the dimensions however, is reinforcement in

Behavioral

Dollard & behavior th concepts of number for such as the recent note two n

Methods the Class

The basic variants (an-penc) Schedule responds daydream unable to S elf-suffi able to c use of the Survey therapy.

patient
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY

significant degree. If these are the findings with respect to amount of social contact, how much greater must be the problem of estimating behaviors more vaguely specified, such as anxiety. Tests used to obtain self-report need to be behavioral, specific, and close in time to the occurrence of the behavior in question.

Behavioral Analysis and Environmental Psychology

Although the reinforcement contingency deals directly with the environment and its effect on behavior (see 220 for a review of this subject with respect to the measurement of abnormal behavior), a separate field of environmental psychology has sprung up in recent years. Its discoveries might well contribute to a behavioral analysis with its methods of objectively describing the “personality” of environments (97, 161, 162). The environment has been described by such dimensions as ecology, organization structure, personality of milieu inhabitants, behavior settings, reinforcement properties of the environment, and psychosocial characteristics. Although only one of the dimensions is conceptualized in terms of behavior theory, behavior theory provides the mechanism through which the dimensions no doubt influence people’s behavior. The more important point here, however, is that the behavior analyst trying to assess the viability of programs of reinforcement or extinction should obtain accurate knowledge about the environment in which the patient resides.

Behavioral Analysis of Traditionally Categorized Disorders

Dollard & Miller (47) provided an early extensive treatment of neurosis in terms of behavior theory. Their analysis, unfortunately, accepted as fact too many of the concepts of psychoanalytic theory. More recently, Salzinger (214) showed how a number of learning paradigms could produce known entities of abnormal behavior such as delusions, depression, and psychophysiological reactions. Because of the rising interest in the depression syndrome, it might be useful to note two new behavioral analyses of depression (55, 160).

Methods of Behavioral Analysis Related to the Classical Conditioning Model

The basic approach to assessment for Wolpe’s behavior therapy and some of its variants consists of the classical psychological procedures, namely the paper-and-pencil tests. Wolpe (286) led the way by using the Willoughby Personality Schedule, a test for neuroticism or unadaptable anxiety reactions. The patient responds to such questions as, “Do you get stage fright?” Other questions deal with daydreams, being easily discouraged, feeling miserable, feeling inferior, and being unable to make up one’s mind. In addition, Wolpe (287) uses the Bernreuter (9) Self-sufficiency Questionnaire primarily to assess the degree to which the patient is able to carry out instructions with regard to self-assertion. He also suggested the use of the Maudsley Personality Inventory. Finally, Wolpe (287) uses the Fear Survey Schedule, which is quite popular in respondent conditioning behavior therapy. This test (288) lists situations that might provoke fear in the patient. The patient indicates degree of disturbance on a five point scale, ranging from “not at
all” to “very much.” The 1969 version lists 87 situations, including noise of vacuum cleaners, falling, sirens, failure, sick people, receiving an injection, birds, lightning, doctors, people with deformities, dirt, and a lull in conversation.

The fear-survey schedules used by Wolpe and others have been fully reviewed by Hersen (94) and by Wolff & Merricks (285). Tasto & Hickson (268) standardized and scaled a 122-item Fear Survey Schedule (FSS). They administered it to a group of 132 male and 107 female undergraduate students at a large university. Standardization of a test presents the therapist with an objective criterion for determining when the fear has been reduced to the level that characterizes most of the population. Standardization on a college population, however, is questionable in terms of how representative it is of the population at large. In a later paper, Tasto, Hickson & Rubin (269) performed a profile analysis for the five factors earlier derived by Rubin et al (210). These factors and their corroboration by profile analysis give some empirical basis for gauging the generalization to be expected from the treatment of one problem to that of another. In 1972, Tasto & Suin (270) administered the same test twice within a 10-week period and found generally high correlations, but not uniformly so, for all factors. More troublesome were the significant decreases in the factor scores from the first to the second test administration. As the authors point out, this “natural” decrease would definitely have to be taken into account when estimating the effectiveness of therapeutic intervention by this test.

Hersen (94) presents additional results on the fear survey schedules. Female subjects, whether undergraduates or patients in inpatient or outpatient therapy situations, have higher fear scores than male counterparts. It is not clear whether this difference is to be ascribed to differences in fears, to the selection of items of the test, or to the difference in tendency to admit to having fears. Normals have the smallest number of fears, neurotics the next larger, and psychotics the largest. Reliability of the fear survey was very high on a split-half basis but not as high on a test-retest basis. Furthermore, schizophrenic subjects showed a significantly lower reliability than normals on a snake fear (.53 as opposed to .79 for normals).

Although the modification of self-assertive responses consists of the conditioning of operant behavior, Wolpe has used it as part of the inhibitory mechanism in his reconceptualization paradigm, and therefore we shall deal with it here. Rathus (202), inspired by the ever greater use of assertiveness training, devised a schedule of 30 items for its measurement. Both test-retest reliability and split-half reliability correlations were above .77. With respect to validity, Rathus had the testers choose a group of people they knew well and whom they could rate on a 17-item rating scale. Inter-correlation of the ratings with the subjects’ responses about themselves yielded a series of positive values, indicating that subjects who appeared to be assertive to friends revealed themselves to be so by their own responses on the self-assertiveness scale. Furthermore, the degree of assertiveness shown by subjects’ stories correlated positively ($r = .70$) to their scores on the assertiveness scale.

It is not surprising that at least some behavior therapists were not satisfied with behavioral diagnoses on the basis of paper-and-pencil tests (69). Rutner & Pear (211) devised a measure of the length of food reinforcement, high from willingness to approach subject to a snake. The closer the subject’s arousal measure a magnitude of the however, specific are more directly

**Methods of Behavior Modification**

Behavioral analysis (77) instructed his own problems as of a particular class.

The single most Kanfer & Saslow present the following situation—is that the also looks for the situation—people this. A motivator now impinging on that might change biological equipment for example, a ground. These I investigated. 5. social relations hindering the comparison of background. K for this type of view but, on the the patient's life. Others, by observation might know about the specific laborator. In their book a patent might therapy that it responses.

Storrow (26}
(211) devised a behavioral measure of the avoidance of a feared object by measuring the length of time a subject is willing to observe a rat pressing a lever on a food reinforcement contingency. This length-of-observation measure distinguished high from low phobic subjects as determined by how close they were willing to approach a rat in a cage. Craighead (41) related the proximity of a subject to a snake to verbal self-report of fear and to level of physiological arousal. The closer the subject was able to approach the snake, the less the physiological arousal measure and the greater the subjective fear measure. Unfortunately, the magnitudes of the correlations were not high. With respect to sexual problems, however, specific physiological measures, such as those of penile volume (4, 285), are more directly relevant.

Methods of Behavioral Analysis Related to the Operant Conditioning Model

Behavioral analysis is an intrinsic aspect of operant conditioning. Goldiamond (77) instructed his clients in behavior therapy; then he told them to think of their own problems as an experiment in which they are trying to modify the probability of a particular class of behavior.

The single most significant effort in the area of behavior analysis stems from Kanfer & Saslow (101), along with a later revision by the same authors (102). They present the following outline for eliciting information: 1. Analysis of the problem situation—is the behavior excessive or deficient? Under this category the therapist also looks for the patient’s behavioral assets. 2. Clarification of the problem situation—people involved in the behaviors and the consequences of the behaviors. 3. A motivational analysis—a listing of the reinforcers, positive and negative, now impinging on the patient, those that could be used for therapy later, and those that might change if therapy is successful. 4. Developmental analysis—patient’s biological equipment and sociocultural background which may limit behavior, as, for example, a limp or blindness, poor education, or different language background. These factors and characteristic behaviors at different times of life are investigated. 5. Analysis of self-control—inadequate or excessive. 6. Analysis of social relationships—the patient’s social networks reveal who can help and who is hindering the patient. 7. Analysis of the social-cultural-physical environment—comparison of the patient’s behavior to that expected by someone of that social background. Kanfer & Saslow explain that the information to be collected for this type of analysis need not all be garnered from the usual patient interview but, on the contrary, ought to be obtained by talking to significant others in the patient’s life, by watching him or her in interaction with those significant others, by observing that patient at work and at play, by interviewing people who might know about the patient, and also by employing paper-and-pencil tests and specific laboratory techniques as they are called for.

In their book, Kanfer & Phillips (100) successfully relate the particular problems a patient might have (as ascertained from the above analysis) to the techniques of therapy that should be used, such as extinction and conditioning of competing responses.

Storrow (260) has devised various check lists for his behavior analysis. The
Personal Data Form, filled out by the patient, has questions such as: “What persons, situations, activities, etc seem to set these symptoms off or make them worse?” (p. 219). He completes the Psychiatric Case Study form on the basis of the Personal Data Form, a physical examination, information from relatives and the patient in the interview, and psychological testing. Although he calls the behaviors he wishes to modify symptoms, he views them as the dependent variables to be modified, rather than the epiphenomena of the medical model. The final part of his form notes his decision of what to change in the patient and by what method. Another variation on the Kanfer-Saslow behavioral analysis is the work by Stuart (261), in which forms are used following the $S^D\ldots S^R$ paradigm. He categorizes what we have called $S^D$s into four classes: 1. instructional stimuli; 2. discriminative stimuli (which Stuart restricts to mean an offer of positive reinforcement for a response); 3. potentiating variables which insure that the reinforcer is effective; (for example, by depriving an individual of those things that will be given later as reinforcers); 4. facilitating stimuli (tools such as books if the subject is expected to read). Pomranz & Goldfried (186) presented another variant of this basic kind of outline. Finally, Carter & Thomas (24) presented 27 categories of “problems in communication” in marital couples. These are for the most part operationally defined and consist of overtalk, fast talk, quiet talk, affective talk, abusive talk, opinion deficit, and topic avoidance, among others.

A very important aspect of behavioral analysis is the use of observation by trained personnel, by the patient, or by a significant other. In many studies it has been shown that the act of observation by interested parties, whether the patient or significant others, increases self-control. In view of the frequent deficit in such self-control initially, it is perhaps not surprising that behavior should improve merely because it has been observed in an objective manner. Salzinger, Feldman & Portnoy (215) trained parents to use a form to record their observations of the behavior of their children by noting the children’s responses, their time of occurrence, the people present at that time, the occasion on which the behavior occurred, and the events that followed that behavior. Typical results showed that the observations were too vague at first, but with shaping through discussion of the recorded observations, the parents became more specific. At that point the parents’ behavior also changed, and occasionally the behavior of the target child improved before any special program was enacted. Apparently the requirement to record what was happening made the parent hesitate long enough at the time of such behaviors as tantrums so as not to reinforce them. In some cases a behavioral change is easy to record objectively, as in the case of weight loss. In most others, unfortunately, despite the therapist’s best intentions, the only measure available is the less easily verifiable patient-recorded behavior. Simkins (234) reported on the lack of data concerning reliability of such records. Electronic devices (227) might eventually yield indices of validity for such observations.

One more type of assessment ought to be mentioned. In 1967, Cautela & Kastenbaum (25) devised a reinforcement survey schedule on which subjects may indicate the extent to which certain items give them “joy or other pleasurable feeling.” Examples are: eating, beverages, nude men, nude women, reading, singing, skin diving, and being praised by the boss for having completed a job well done. Thorndike’s weeks ranged from at least some of the

Conclusion

There are now a include the usual psychology. They focused and in req obtained. Much m have well-standard are some specific t

2. THE MEAS OF DEPRESSI

Joseph L. Fleiss

After long negle depression resum on many fronts: clinical. Even tho the individual deq in general. This n must be applied t nondepressives a measure change) (in order to asc communicate on

This review w and classify dep unipolar depre: previous episcol unipolar depress been reviewed treatment of de characterize deq usual activities, suicidal thinkin

2 The author is York City.

The preparat National Institute Sallone of the Ds and Dr. Alexand York State Psych
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY

Thorndike & Kleinknecht (272) found that test-retest reliability over 5 weeks ranged from .40 to .91 for various clusters of reinforcers. It is suggested that at least some of the clusters are sufficiently reliable for therapy research.

Conclusion

There are now a number of methods of behavioral analysis available. They include the usual techniques for obtaining information in the area of clinical psychology. They differ in the main from the older techniques in being more focused and in requiring fewer assumptions about the meaning of the responses obtained. Much more work needs to be done, however, before we can say that we have well-standardized methods of doing behavioral analysis. What we have so far are some specific tests and some general outlines of inquiry.

2. THE MEASUREMENT AND CLASSIFICATION OF DEPRESSION

Joseph L. Fleiss

After long neglect at the expense of research in schizophrenia, research in depression resumed actively in the early and middle 1960s. The research has been on many fronts: biochemical (224), epidemiological (131, 233), genetic (181), and clinical. Even though the unit of analysis in these research endeavors is invariably the individual depressed patient, the results are meant to be applied to depressives in general. This means that in each of these kinds of research, evaluation methods must be applied to measure a subject's degree of depression, in order to screen out nondepressives as not suitable for study, to describe one's resulting sample, or to measure change, and to classify him into one of a number of diagnostic categories (in order to ascertain which kinds of patients one's results apply to, and to communicate one's results to others).

This review will be restricted to methods that have been employed to measure and classify depression. Attention will be limited as far as possible to so-called unipolar depression, i.e. depression as the primary current complaint with no previous episodes of mania or hypomania. Research into the distinction between unipolar depression and bipolar depression (i.e. manic-depressive illness) has been reviewed elsewhere (2, 283), as has research into the etiology and treatment of depression (39, 258). Some of the signs and symptoms used to characterize depression are, in addition to a depressed mood, a loss of interest in usual activities, an inability to concentrate, psychomotor retardation or agitation, suicidal thinking, weight loss, and feelings of hopelessness.

2 The author is also affiliated with the Division of Biostatistics, Columbia University, New York City.

The preparation of this review was supported in part by grant MH 23964 from the National Institute of Mental Health. Criticisms provided by Drs. David Dunner and Frank Stallone of the Department of Internal Medicine, Dr. Jean Endicott of Biometrics Research, and Dr. Alexander Glassman of the Department of Biological Psychiatry (all at the New York State Psychiatric Institute) are gratefully acknowledged.
The Measurement of Depression

Rating Scales. Rating scales most frequently employed for measuring depression are of three kinds: self-report questionnaires to be completed by the subject; scales to be filled out by a professional after interviewing and observing the subject; and scales to be completed by a nurse or attendant on the basis of naturalistic ward observations.

Self-report scales for depression have been developed by Beck et al. (6), Pilowsky & Spalding (184), Rockliff (206), the Wakefield Hospital group (241), and Zung (296). The major advantage to relying on the subject to rate his own condition is the saving of professional time. Self-report scales are thus economical sources of data for large-scale surveys. They have been shown to be sensitive in assessing a patient’s degree of recovery from a depressive episode (190, 198), but not in estimating the severity of an acute episode experienced by a psychiatric inpatient (190). Their usefulness may be greater with outpatients and private patients.

One disadvantage to self-report scales is that some subjects may be unable to understand the intent of the questions. Because of denial or deliberate falsification, the severity of depression is frequently underestimated by the subject himself (178, 206, 275). In addition, many newly admitted inpatients are too acutely ill even to complete a short questionnaire within a few days of admission (250).

An alternative procedure for measuring depression is a clinical interview of the subject followed by or coincident with the completion of a rating scale. Scales for depression to be administered by clinicians or other professionals after an interview have been developed by Hamilton (86), Pilowsky and his colleagues (184), and Zung (298). Wechsler et al. (277) developed a scale which relies on both sources, the patient’s self-report as well as a professional’s judgments.

Spitzer & Endicott (248) have indicated how the interview is superior to other methods of data collection in psychopathology. They point out, however, that differences among interviewers in the topics they cover and in how they phrase their questions, and differences for the same interviewer in how he interacts with different subjects, can greatly affect the final ratings. Thus sizable proportions of the variance in scores from rating scales based on an interview may be associated with differences among interviewers rather than differences among subjects (58).

In order to reduce this extraneous variation, structure may be imposed on the interview by specifying for the interviewer the order of topics to be covered, the questions to be asked, or even the probes to be put to the subject. Only Zung’s later scale (298), among those specifically designed for depression, seems to incorporate a structured interview.

The self-report and interviewer-completed scales cited above limit the measurement of depression to that currently being experienced by a subject, and are thus useful for sample description and for establishing baselines in the study of change. They do not cover previous episodes or other aspects of psychiatric history. As pointed out by increasing numbers of investigators (3, 54, 205), however, history is at least as important as current state for describing depression when one’s purpose is classified developed by Spitzer & both current and past a and schizophrenic beha. Ratings made by nu provide a useful adjunct even reveal patterns of Raskin and his colleg nurses’ ratings of the w: either self-report or in transitory manic states the interviewers were a.

A limitation on nat applied only to hospit community. In additio pathology (e.g. emotio ratings (199).

Each source of int nurse—provides data a in conflict. The multi (199–201) represent a portrait of depression.

Although most of f currently experienced nonpatients (296, 298 course of illness (18) specify the kinds of whether the data are

Sleep Disturbance I disturbance in depres early morning wake neurotic and psych confirmed this classi of neurotic and psy waking was one of characterizing psych.

Other studies (22). A major source of d information. Some quality of his sleep (by nurses throughout negative, reporting pressives.

Two other studie
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY

For measuring depression, completed by the subject, rating and observing the individual on the basis of

Bond and Beall (6), a modification of the Hospital group (241), have developed the RLA report to rate his own behavior. Results are thus economical and have been shown to be sensitive in detecting depression (190, 199), but more accurately measured by a psychiatric scale in patients and private practice patients. The RLA scales may be used to identify the presence of depression, either by the subject himself or by the observer. The scales are administered by inpatients and outpatients for the first few days of admission for the diagnosis of depression.

Psychiatric interview or rating scale. Scales for psychiatric professionals after an examination of patients and his colleagues (190) rely on both observations and self-reports. The interview is superior to other methods of assessment, although, however, that it is important not to measure the way they phrase the questions. The psychiatrist interacts with the patient, and the interview is designed for different populations in different settings.

R. C. L. Zung (199) was one of the first to use the interview as a tool for assessing depression. Only Zung's later scales seem to incorporate the variables into the rating system. The interview scales limit the mean rating for a subject, and are useful in the study of the effects of psychiatric treatments (3, 54, 205). The interaction between the interviewer and the patient is the key to the development of depression where the patient might not mention it.

One's purpose is classification. Only one scale, a structured interview procedure developed by Spitzer & Endicott (249), appears to exist for covering thoroughly both current and past aspects of depression. The scale also has sections for manic and schizophrenic behavior.

Ratings made by nurses or other ward personnel of patients' overt behavior provide a useful adjunct to ratings based on an interview (18, 198, 199). They may even reveal patterns of behavior that are not picked up by other rating methods. Raskin and his colleagues (200, 201), for example, found a factor of excitement in nurses' ratings of the ward behavior of depressed patients which was not present in either self-report or interview-based ratings. The nurses may have noted either transitory manic states or periods of extreme agitation that neither the patients nor the interviewers were aware of.

A limitation on naturalistic observations of ward behavior is that they may be applied only to hospitalized patients and not to outpatients or nonpatients in the community. In addition, ward personnel may be insensitive to aspects of psychopathology (e.g., emotional withdrawal) which are revealed by the other scales of ratings (199).

Each source of information—the patient himself, the interviewer, and the nurse—provides data about various aspects of depression, but these data are often in conflict. The multitrait-multimethod studies of Raskin and his colleagues (199-201) represent an attempt to dovetail these observations into a coherent portrait of depression.

Although most of the scales cited above are intended to describe the depression currently experienced by psychiatric patients, some were designed for use with nonpatients (296, 298), and at least one was especially designed for describing course of illness (18). Before selecting a scale, therefore, the investigator must specify the kind of patient he intends to study, who is to make the ratings, and whether the data are to be analyzed cross-sectionaly or longitudinally.

SLEEP DISTURBANCE IN DEPRESSION. A stereotype characterizes the place of sleep disturbance in depression. Neurotic depressives are supposed to experience difficulty in falling asleep, whereas psychotic depressives are supposed to experience early morning wakening. See below for a review of the distinction between neurotic and psychotic depression. A number of studies (85, 117, 209) have confirmed this classic picture. In a review of six studies contrasting the symptoms of neurotic and psychotic depression, Costello (38) found that early morning wakening was one of only two signs agreed upon by at least four of the studies as characterizing psychotic depression.

Other studies (22, 40, 150, 153), however, have failed to confirm this stereotype. A major source of discrepancy between studies is a reliance on different kinds of information. Some studies have relied on the patient's own recollection of the quality of his sleep (22, 85, 117), and others have relied on the observations made by nurses throughout the night (40, 150). The latter kind of study has tended to be negative, reporting small if any differences between neurotic and psychotic depressives.

Two other studies (168, 185) have relied on both the patient and the nurse as
sources of data, but their results disagree. One (168) reported good agreement between nurses' observations and patients' reports of the quality of sleep, and the other (185) that patients tend to report their sleep to be poorer than nurses perceive it. Weiss and his colleagues (278) found that neither nurses nor patients were able to agree well with information provided by electroencephalographic studies.

The definitive description of the patterns of sleep disturbance in depression will only come from all-night EEG studies. Kales & Berger (99) reviewed the relatively small number of such studies performed through 1969. Their review suggests that neurotic and psychotic depressives have similar patterns of sleep disturbance, with the disturbances perhaps being more severe in the psychotic than in the neurotic group.

RATER PERCEPTIONS Even if structure is imposed on the clinical interview, which is the most important source of information about depression, extraneous variation will remain due to differences in how raters perceive what the subject reports and how he behaves (58). In one study (184), for example, ratings of depressed patients made by medical students relatively unsophisticated in psychopathology correlated much more strongly with patients' self-reports \( r = .91 \) than did ratings made by either experienced psychologists \( r = .40 \) or psychiatrists \( r = .36 \). The higher correlation achieved by the students may be attributed either to their being less influenced than the professionals by theoretical formulations, or to their naively accepting as fact, without further probing, whatever the patients reported.

Reliance on standardized and structured procedures for assessing depression assures a reduction in, although not an elimination of, both bias (which might result in spurious differences) and unreliability (which might obscure real differences) (108). Such reductions are especially likely in the context of a single study, for the raters can be trained in the comparable use of such procedures.

The theoretical background of the rater is known, however, to affect strongly how he perceives and rates depressive symptoms (38, 51, 107, 108, 156, 179). No amount of standardization and structuring, therefore, will guarantee comparability across studies if the personnel employed in each perceive these symptoms differently.

The Classification of Depression

RELIABILITY OF DIAGNOSIS The major diagnostic subdivisions of the unipolar depressive disorders are neurotic depression, psychotic depression, and involutional melancholia (1, 74). On the basis of evidence reviewed by Zubin (291), as well as more recent data (37), one must conclude that the routine diagnosis of depression is grossly unreliable.

In the original reports, data on diagnostic agreement were presented as simple proportions of agreement, uncorrected for the degree of agreement expected by chance alone. Because chance-expected agreement varies across studies as a function of the base rates with which the diagnoses are made, the uncorrected proportion...
proportions are not comparable. A statistic, kappa, has been developed for measuring how well two (32) or more (59) judges agree beyond the level predicted by chance.

Values of kappa less than or equal to zero imply that agreement is no better than that predicted by chance, and values greater than zero that agreement is better than that predicted by chance. Kappa assumes its maximum value of +1 only when agreement is perfect, and is interpretable as an intraclass correlation coefficient (61, 127).

Three studies (7, 37, 221) report sufficient data on the reliability of the diagnosis of depression to permit the calculation of kappa. For neurotic depression, kappa varies from .10 to .47; for psychotic depression, from .19 to .30; and for involutional melancholia, from .21 to .38. When the subcategories are combined into the single diagnosis of depression, the resulting mean value of kappa is only in the neighborhood of .30.

Serious cross-national differences exist in the criteria for diagnosing depression (37, 110, 230), with British psychiatrists tending to diagnose depressive disorders relatively more frequently than American psychiatrists. Equally serious are the biases against diagnosing depression in lower class patients in general (131) and blacks in particular (98, 144, 235). When blacks and whites are assessed by means of standardized procedures, the difference in the prevalence rate of depression all but vanishes (236, 273, 297).

A study of New York hospital psychiatrists (83) revealed that they were sensitive to the depression exhibited by some of their patients, but were reluctant to bring this sensitivity to bear in making their diagnoses. A sample of patients was identified on the basis of a standardized interview conducted by independent investigators as suffering from severe depression, but with no signs of schizophrenia. These patients were treated with antidepressive therapy by the hospital psychiatrists significantly more frequently than were other kinds of patients, but were nevertheless more frequently diagnosed schizophrenic than depressive.

The inability of clinicians to agree well on the diagnosis of depression, the systematic bias against diagnosing depression in certain kinds of patients, and the tendency among American psychiatrists to underdiagnose depression must be borne in mind as major reasons for disagreements among the studies to be reviewed below.

THE DICHOTOMY BETWEEN NEUROTIC AND PSYCHOTIC DEPRESSION Of the three major subdivisions of the unipolar depressive disorders, involutional melancholia stands out as apparently not meriting status as a separate diagnostic entity. With the trivial exception of age, few differences of practical importance in etiology, clinical characteristics, or outcome have been reported between involutional and other depressives (2, 43, 106, 155).

The validity of separating the other two subdivisions, neurotic depression and psychotic depression, however, has been the subject of intensive research and debate. An alternative dichotomy, reactive versus endogenous depression, over-
laps the neurotic-psychotic dichotomy to such an extent (5) that neurotic depression and reactive depression are virtually synonymous, as are psychotic depression and endogenous depression.

Reactive or neurotic depression is presumably characterized by the existence of a precipitating event, anxiety in the clinical picture, difficulty in falling asleep, and poor premorbid adjustment. Endogenous or psychotic depression is presumably characterized by the absence of a precipitating event, retardation, early morning waking, and a history of previous depressive episodes. Personality variables appear not to be useful in distinguishing between the two types (68), nor perhaps is the existence of life-changing events prior to the onset of illness (271).

Some investigators have either suggested (105, 154, 156) or reported data which suggest (116, 174) that psychotic or endogenous depression is a tightly organized syndrome representing a well-defined illness, whereas neurotic depression is a term applied to a heterogeneous collection of signs and symptoms frequently found together with other neurotic features, personality disturbances, and hysteria.

The validity of this view has been brought into question by Costello's review (38) of six studies (22, 87, 117, 208, 209, plus an unpublished report) which factorially separated features of neurotic from features of psychotic depression. He found that the studies tended to agree on the variables loading on the factor for neurotic depression, but did not agree on the variables loading on the factor for psychotic depression. Only psychomotor retardation was reported by all six studies as a symptom loading high on psychotic depression. Early morning waking was cited as an important feature by four studies, and all other features were cited by three or fewer studies.

The overwhelming importance of retardation in distinguishing between different kinds of depressed patients has been confirmed by other studies (62, 170). It is crucial, however, that psychomotor retardation be measured separately from flat affect (62), something not possible with all rating scales (142, 256).

In addition to the studies reviewed by Costello (38), others (104, 116) have found that ratings of depressed patients factor either into a bipolar factor, with features of the two kinds of depression loading on opposite ends, or into two factors, one defined by signs of neurotic and the other by signs of psychotic depression. The inference frequently drawn from such results has been to confirm the validity of the diagnostic separation between neurotic and psychotic depression. Such an inference is erroneous, for factor analysis is not by itself informative about the existence of types (51, 108, 120, 163, 207).

One criticism leveled at the studies reporting bipolar factors has been that bipolarity may result from raters' perceiving the two kinds of depression as distinct, and thus tending to rate patients as exhibiting either signs appropriate to one type or signs appropriate to the other, but rarely signs appropriate to both (51, 107, 156). When factor analyses have been applied to ratings made by professionals not committed to the dichotomy between the two kinds of depression, neither bipolar factors nor separate factors for neurotic and psychotic depression have been found (148, 165, 199).

More germane is an examination of patients on whom have been examined discriminant function analysis of depressive reported bimodal 111, 170, 179, 20, 116). The same t distributions, anxiety.

Another reason is that the results are equivocal results subtypes. Unless nearly equal pain distribution (67), only visually, or but more complications fits a given bution (60, 284), sufficiently many by the central lit

**NUMERICAL TYPE**

frequency distril (80, 173, 174, 18 163) to cluster de some of these n members of differntially on deficiencies in a yield artificial ment be litt treatment and o effecting the app

One of the ap by Paykel (174), and the others rt in interpreting t different source wards of mental social and cultu the patterns of f therefore, that d from the variou
More germane to the validation of a typology than the results of factor analyses is an examination for clustering or multimodality of the frequency distributions of patients on quantitative variables (87, 163). Increasing numbers of investigators have been examining and reporting distributions of both factor scores and discriminant function scores best distinguishing neurotic depressives from psychotic depressives, but the results have been contradictory. Some studies have reported bimodality (22, 53, 73, 222), others have reported unimodality (105, 106, 111, 170, 179, 207, 271, 294), and yet others have reported equivocal results (104, 116). The same biases which give rise to bipolar factors may give rise to bimodal distributions, and such biases may explain some of the positive findings of bimodality.

Another reason—a statistical one—may account for some of the negative and equivocal results. Multimodality is not a necessary condition for the existence of subtypes. Unless two underlying distributions are well separated or are present in nearly equal proportions, their mixture will result in a unimodal but skewed distribution (67). It is therefore inappropriate to examine a frequency distribution only visually, or to apply only goodness-of-fit tests for normality. A more powerful but more complicated procedure is to test whether a mixture of normal distributions fits a given distribution significantly better than does a single normal distribution (60, 284). If the factor or discriminant function scores are weighted sums of sufficiently many variables, the model of underlying normality becomes tenable by the central limit theorem (60).

Numerical typologies of depression. Rather than looking for modes in a frequency distribution as evidence for the existence of types, other investigators (80, 173, 174, 182, 183) have applied methods of numerical taxonomy (70, 141, 163) to cluster depressed patients directly into homogeneous subgroups. Although some of these numerical typologies have been validated by demonstrations that members of different types respond differentially to treatment (96, 175) or perform differentially on other tests (176, 184), the results must be suspect because of deficiencies in most clustering methods (50, 63, 67), especially their tendency to yield artificial types. Thus the demonstration of differential response to treatment may be little more than a demonstration of a correlation between response to treatment and one or more of the variables (e.g., retardation) most important in effecting the apparent separation between patients (60, 170).

One of the apparently more promising typologies of depression is that derived by Paykel (174). He found four types, one corresponding to psychotic depression and the others representing subdivisions of neurotic depression. A major difficulty in interpreting these types is that Paykel's sample of depressives was selected from different sources: outpatient clinics, emergency services, day care centers, and the wards of mental hospitals. The importance has been recognized of the effects of social and cultural variables on the type of treatment patients seek (121) and on the patterns of psychopathology they present (103, 131). It was hardly surprising, therefore, that differences in the patterns of depression were found among patients from the various sources of Paykel's sample (120, 177), and therefore that his four
types overlapped considerably with the sources of his sample (176). One cannot tell, therefore, whether his typology is one of depressed patients or of facilities treating depressed patients.

Aside from the defect in Paykel's sample selection, his work represents a paradigm for future research. Standardized observations of mental state and history should be made by professionals not committed to any theoretical model of depression, and should be used according to specified rules (205) to select subjects whose primary disorder is depression.

The resulting data should be subjected to the most powerful analytic methods available. The Friedman-Rubin clustering method (70), the one used by Paykel (174), seems to be one of the best available. Within any study, however, one's sample of diagnosed depressives must be derived from one source only (e.g. public mental hospitals or a probability sample of a community) rather than from many sources. It is better to describe one particular kind of depressed patient precisely than to describe many kinds of depressed patients ambiguously.

3. PSYCHIATRIC EPIDEMIOLOGY

Barry Gurland

The generally healthy state of psychiatric epidemiology is evidenced by the tone of the review papers that have recently appeared on this subject (34, 36, 81, 123), by the scientific rigor of the papers in the international conference in Aberdeen (90), and by the appearance of the first textbook devoted exclusively to this topic (35).

The Symposium on Psychiatric Epidemiology, held at Aberdeen University in July 1969, was the first of its kind in nearly a decade and produced almost 20 papers of high standard, published as a volume in 1970 (90). Gruenberg (81), looking back to the pioneering Milbank Memorial Fund meeting on psychiatric epidemiology in 1949, and forward to developments in the near future, concludes that "the epidemiology of mental disorders has become an established category of research in the related scientific disciplines," and he regards it as now inappropriate, for epidemiological purposes, to view mental disorder in isolation rather than as one of the consequences of a variety of conditions, since the latter view provides a more useful unifying theme. Thus, in a very complimentary to the Aberdeen proceedings, he sees no further necessity for such meetings. Although that may be true for substantive studies, the methodological problems of assessing mental disorder seem sufficiently peculiar to this class of disorders to warrant its delineation as a special field for some time to come.

Among the major recent advances in the methodology of psychiatric epidemiology has been progress towards the development of reliable and valid measures for assessing psychopathology in different national or cultural groups and across language barriers; for recording the interaction between patients and treatment services by means of case registers or record linkages; and for assessing forces in the biological or cultural milieu which might show associations with psychopathology.
As a measure of psychopathology, psychiatric diagnosis has been much improved by the use of semistructured mental state interviews to standardize the scope and style of collecting information for diagnosis and to allow an analysis and synthesis of the discrete items of psychopathology which relate to diagnosis; and by the use of glossaries to aid in the definition of diagnostic labels. The use of such standard diagnoses and diagnostically oriented symptom analysis for epidemiological purposes has been demonstrated by the US-UK Diagnostic Project in a study of recent admissions to public mental hospitals in London and New York (37, 84). Diagnoses and symptom ratings made by the project’s research psychiatrists were shown to be cross-nationally reliable, whereas the routine diagnoses by hospital psychiatrists were based on different though overlapping criteria in the two cities. The New York hospital psychiatrists had a much wider concept of schizophrenia compared to their London colleagues so that the hospital statistics gave a misleading picture of the relative distribution of diagnostic groups in the two cities.

Unfortunately there is no evidence as yet that semistructured interview techniques and glossaries of diagnostic terms have done much to improve the reliability of routine hospital diagnoses which are potentially the backbone of epidemiological work based on admission records or case registers. Kendall (109) found that there was no appreciable change in the specification of diagnostic labels by psychiatrists in English mental hospitals after the promulgation of an official glossary for diagnostic categories in the mental disorders section of the International Classification of Diseases.

This is not to beg the question of whether psychiatric diagnosis is necessarily the best reference point for epidemiological work. Some general population studies rely on inventories of symptoms which, though highly associated with psychiatric impairment, are not intended to cover the wide range of symptoms essential to the diagnostic process. However, when contrasting results are found between surveys based on diagnosis or diagnostic constellations of symptoms and those based on single symptom rates alone, then the most serious questions arise about the relative usefulness of these contrasting measures of psychopathology for the particular purpose of that survey. Examples of such a contrast may be found in studies on the age distribution of depression where diagnostic surveys show the well-known bell-shaped curve of prevalence between 20 years and 60 years of age, while surveys based on symptom counts may show an age-prevalence curve resembling an inverted bell (71, 226). In terms of the purposes of psychiatric epidemiology (which have been described as serving the interests of public health efforts, clinical recognition of treatable cases, and research into the etiology of psychopathology) it might seem that both diagnostic and symptom orientations are worth exploring, as well as other factors in case definition. Clearly a key area in epidemiological work is the identification of the intervening variables between symptoms and clinical diagnosis in a nonpatient population.

Dohrenwend et al (44) have directed their attention to what they consider as the “central unsolved problem of psychiatric epidemiology—the problem of how to conceptualize and measure psychiatric disorder independently of treatment status.” They collected data by means of two different interview techniques, one a
mental state interview with open-ended questions and the assignment of responses to defined categories by rater judgment, and the other a screening interview for psychiatric impairment with prescribed response alternatives presented to the respondent. The interviews were conducted by psychiatrists on psychiatric patients, community leaders, and a probability sample of five ethnic groups in a community. The investigators found, *inter alia*, that relations between judged disorder, ethnic status, and class status varied with the type of research interview and whether judgments were made face to face or by “blind” review of the protocols. Community subjects were judged to be more ill on the basis of protocol review than at face-to-face interview, and this discordance was greatest for community leaders. The weighing of positive strengths that are observed in a face-to-face situation but not recorded is postulated as contributing to the discordance between interviewer-reviewer ratings of psychiatric disorder.

**Assessment Across Racial Boundaries**

The study cited above is a fine example of scientific carelessness and restraint in that the methodology is being painstakingly worked out before the substantive goal is tackled (in this instance, the relationship between social class and psychiatric disorder). This is not always the case. Often the substantive cart is put before the methodological horse, or methodological contributions are the result of secondary analyses of data collected for some other purpose. Thus it is not surprising that there have been many unconvincing statements and theoretical inferences in the literature about the differences between ethnic or racial groups in the frequency and forms of mental disorder. For example, it has been claimed that blacks, compared with whites, less often suffer from depressive disorders and more often suffer from schizophrenia; when depressed, the blacks are less likely to show guilt or suicidal trends and more likely to complain of somatic symptoms; and when schizophrenic they are more likely to show hallucinations. However, as Simon et al (236) remarked, “a major drawback to almost all the studies [until recently] is that, for data on psychopathology or diagnosis, they relied on hospital statistics, on the patients’ hospital records, or, in the few instances in which patients were actually interviewed, on the psychiatrists’ clinical impression.”

Several recent papers have greatly improved the methodology in the assessment of psychopathology across racial boundaries. Simon et al (236) reanalyzed data from the US-UK Diagnostic Project in which semistructured mental state interviews were carried out by research psychiatrists on recently admitted patients in public mental hospitals in London and New York. When the type and degree of psychopathology was controlled, a bias was revealed in the hospital diagnosis towards schizophrenia rather than depressive disorder in blacks compared to whites. When the research diagnosis was controlled there were no significant interracial differences in symptomatology among schizophrenics, but the black depressives reported significantly more anxiety, somatic complaints, and irritability than did the white depressives.

Tonks et al (273) reanalyzed data from a study on depressed inpatients and outpatients in interviews and demonstrated the difference among patients who were not diagnosed and patients who were diagnosed. The white group, on the other hand, showed no such differences. A significant interaction was noted, with patients who were not diagnosed and patients who were diagnosed, with the white group showing no significant differences. The black group, on the other hand, showed a significant interaction, with patients who were not diagnosed and patients who were diagnosed, with the black group showing no significant differences.

Leff (129) has suggested that the difference in symptomatology among schizophrenics who are not diagnosed and schizophrenics who are diagnosed is due to the different social and cultural backgrounds of the patients. It is not clear, however, whether the differences in symptomatology are due to the diagnostic criteria or to the different social and cultural backgrounds of the patients.
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY

outpatients in New Haven, all of whom received semistructured mental state interviews and a diagnosis of depressive disorder from a research psychiatrist. Significant interracial differences were found for several symptoms, but were much reduced when social class was controlled, and all but vanished when patients were matched for severity of illness; only the symptom of feeling hopeless remained significantly different between the races (more frequent in whites than blacks). As a whole, the black group was less severely depressed than the white group, possibly because the black patients tended to seek treatment earlier in the course of their disorder than did the whites.

Liss et al (140) and Welner et al (279) examined patients who had been admitted to a private hospital in St. Louis, Missouri. One hundred and nine of these patients were followed up and given a semistructured mental state interview an average of 3 years later. Blacks, compared to whites, were found significantly more often to show delusions (of reference, bodily change, and grandeur) and hallucinations (both auditory and visual). These differences were regarded as characteristic of the racial groups rather than as diagnostic factors, partly because symptoms and diagnosis were more highly correlated in whites than in blacks.

Leff (129) has provided a closely argued hypothesis that as cultures develop the elements of the culture learn to express emotional states in a more differentiated form. Primitive expression of a limited variety of somatic experiences grows into sophisticated references to a wide range of emotional states based upon learning of the different social contexts in which the somatic experience can occur. He suggests that possibly this progression of emotional differentiation parallels the development of individual as opposed to group identity. In support of his basic hypothesis, Leff reanalyzed data from the International Pilot Study of Schizophrenia (289), in which a semistructured mental state interview was used for psychiatric evaluation in five “developed” and four “developing” countries, in each case in the appropriate indigenous language. The emotions of anxiety, depression, and irritability showed generally higher correlations (i.e. less differentiation) among patients within developing than within developed countries. Leff also reanalyzed data from the US-UK Diagnostic Project (New York sample) to show that among public mental hospital admissions blacks showed less differentiation than whites even when the patients were matched for social class. In a further data analysis, based on the reliability trials in the IPSS study, Leff (130) showed that his previous findings on the differentiation of emotional states were probably not an artifact of differences in the rating behavior of psychiatrists from developing and from developed countries.

It would appear that advances in method are enabling statements and theories on interracial differences in psychopathology to become more precise but no less challenging or conflicting than previously. It remains unclear whether conflicts between different studies result, on the one hand, from differences in methods of data collection, in rater bias, selection of the samples (e.g. in the referral mechanisms for becoming a patient), diagnostic criteria, or, on the other hand, from cultural differences between the samples. However, an important product of this work has been to underline the importance of semistructured interviewing and
strict criteria for diagnosis as a means of reducing diagnostic errors in black patients, particularly with regard to detecting affective disorders.

Effects of Language on Assessment

The study of the relationship between race and psychopathology often confounds the effects of language and culture. By way of contrast, the studies of Marcos et al. (146, 147) demonstrate the care required in experimental design to enable the isolation of the effects of interview language on the evaluation of psychopathology. These investigators prerecorded on audiotape the questions in a mental state interview and replayed the tapes to ten bilingual psychiatric inpatients with a hospital diagnosis of schizophrenia. Half the initial interviews were in Spanish and half in English, with a crossover design for a second interview within 24 hours. The patients' responses were videotaped and rated by English and by Spanish speaking psychiatrists. Higher levels of pathology were rated from the English than from the Spanish interview, even in those patients who had higher vocabulary scores in English than in Spanish. In a subsequent study, bilingual raters reviewed the videotapes and directly compared Spanish and English responses to the same questions. They noted several possible reasons for the higher scores in the English than in the Spanish interview, including frank contradictions (i.e., admits to pathology in English but not Spanish interview), increased tenseness and uncommunicativeness in English, downgrading of somatic complaints by Spanish psychiatrists because of high frequency in the Spanish culture, and the incoherent structure of the patients' speech in English resembling the features of depression (slowness, pauses, etc.). However, having said all that, the authors attempt to resolve the conflict between their results and those of some previous workers by speculating on the interaction between language and the social context of the interview. Even this most careful study, because of its limited scope, leaves the generality of its findings in great uncertainty.

Assuming that the psychopathology shown by a patient varies with the language in which he expresses his symptoms, there arises the question of whether the patient is more or less "sick" in one or the other language. Careful discussions of the different concepts and components of sickness have recently been given by Wing et al. (281) and Mechanic (151). However, for the purposes of this review it is important to emphasize that the interview assessment is at best only a predictor of the patient's behavior under the usual circumstances of his life, of his response to treatment, of his future progress, and possibly of the etiology of his disorder. There would seem to be the appropriate criteria to which we must ultimately refer to establishing the relative value or validity of cross-cultural variations in symptoms.

Fortunately, there now exist translations from English into various foreign languages for most of the well-established techniques for reliably assessing psychopathology. For instance, the Present State Examination has been translated into Chinese, Czech, Danish, Hindi, Russian, Spanish, and Yoruba (289); the Psychiatric Status Schedule has been translated into Polish, Spanish, Italian, Hebrew, and Flemish (R. L. Spitzer & J. Endicott, personal communication, 1974). For the most part comparability between the different language versions has been established by independent "back-translation." An example of a fully designed and cross-Cultural A. Study on the Effect of German versions of psychopathology questions on the identification of language groups is presented in a book by Manaster.

Despite the many solid ground is evident, the categorizations may be remodeled with the state interview an of Diseases. The by analysis of various diagnoses of schizophrenia were then subjected to the analysis between symptoms or signs of schizophrenia, the presence of four diagnosed schizophrenia. The case was about 65% promising results be reliably rated accurately identified purposes than of mental response, etiological factors included in cases of cross-sectional sampling and diff
fully designed and organized multilingual study is the International Collaborative Study on the Efficacy of Loxapine, involving the use of English, French, and German versions of the Structured Clinical Interview (18a) for the assessment of psychopathology at initial and follow-up interviews. Such studies may contribute much to the identification of the predictive equivalence of symptoms between language groups. A comprehensive account of the problems and provisional solutions in cross-national social-psychological research is provided in a recent book by Manaster & Havighurst (145).

Cross-Cultural Assessments

Despite the many difficulties in assessing psychopathology cross-culturally, some solid ground is evident in that the symptom profiles of certain diagnostic categories may be remarkably similar between different national, racial, and cultural groups. From this solid ground suggestions have been made for standardizing diagnoses by nominating a limited set of symptoms to define a group of patients whose diagnosis seems to meet with universal agreement, or by utilizing statistical techniques for reliably replicating the classification achieved by clinical diagnosis.

Carpenter et al (23), reporting on data from the International Pilot Study of Schizophrenia collected by psychiatrists from centers in the nine participating countries, propose the use of 12 signs and symptoms of psychopathology as an operational definition of schizophrenia. In this study, 1202 patients recently hospitalized with evidence of psychosis were examined with a semistructured mental state interview and given a diagnosis according to the International Classification of Diseases. The wide range of items in the interview schedule were winnowed out by analysis of variance to find the strongest discriminators between cases diagnosed schizophrenia and those given some other diagnosis. These discriminators were then subjected to a stepwise discriminant function analysis using the distinction between schizophrenia and other diagnoses as the criterion. Twelve symptoms or signs were thus selected for differential diagnosis. The more of these symptoms that were present, the more likely was the case to be diagnosed schizophrenia. Taking into account both the initial and the replication samples, the presence of four or more of these symptoms identified correctly over 90% of the diagnosed schizophrenics and included less than 40% of the cases receiving other diagnoses. The corresponding figures for the presence of six or more symptoms was about 65% correctly and 5% incorrectly predicted to be schizophrenic. These promising results invite further work to establish whether or not the key items can be reliably rated by other research groups; whether the group of schizophrenics accurately identified in this way are more or less useful for research or clinical purposes than other groups of schizophrenics (e.g. for predicting outcome, treatment response, biological indices, genetic findings, and associations with possibly etiological factors); and whether nonschizophrenics can be similarly accurately identified for control groups in schizophrenia research.

Use of Case Registers

If cross-sectional epidemiological studies are difficult enough, an even more complex and difficult task is the collection of data on a given population over time.
This aspect of epidemiology is vital for the description of the long-term course of mental disorders, for chronicling utilization of services, for assessing incidence rates, for studying differences between generations (successive cohorts), and for relating etiological factors to consequences remote in time. Case registers and cohort studies are being used with growing sophistication for such longitudinal epidemiological inquiries.

Richman (204) has recently reviewed the use of case registers of psychiatric care in epidemiological research. He defined case registers as "systems whereby records from a specified set of psychiatric facilities are collected for individual persons from a defined population and accumulated over time," and noted that the usefulness of the register increased with the duration of its existence. As of now there are several case registers in English-speaking countries which have been in operation over a decade, including those in Aberdeen, Scotland; Duchess County and Rochester in New York; and Camberwell, London. However, many registers have been born and abandoned in the past 15 years, largely because of their expense and because the effort sunk into data collection was not matched by an equal devotion to data analysis and practical utilization of the results. It is therefore particularly encouraging that Wing and his colleagues (281) have demonstrated the rich potential of a register in a volume which reports the use of the Camberwell Psychiatric Register between 1964-71 for monitoring and planning the introduction of a community psychiatric service. Wing emphasizes that his register is a scientific instrument "to measure the extent and type of current use of services, to facilitate the examination of patterns of contact over time, to monitor changes accidentally or deliberately introduced, to allow estimation of future trends, to indicate researchable issues and to act as a sampling frame for more intensive studies."

Grenenberg (81) draws attention to the importance of "population laboratories" enabling longitudinal studies on cohorts of births, relating the course of the pregnancy to the child's mental development. He refers to the birth registers of cohort studies in Birmingham, Newcastle, Cardiff, and Aberdeen.

Although the Scandinavian countries have long since shown the value of a national network of records on mental health statistics, similar enterprise has been relatively recent in the United States and the United Kingdom. In the United States, Kramer, Chief of the Biometry Branch, National Institute of Mental Health, and his colleagues have organized and analyzed the national statistics on admission rates to mental hospitals and other relevant inpatient and outpatient facilities for the purpose, among others, of applying epidemiological methods to the planning and evaluation of community mental health programs. Their recent work has emphasized the importance of taking household size and composition into account in identifying groups at high risk for requiring psychiatric treatment (123); of age variation in the pattern of use of various types of psychiatric facilities (124); and of the need for improving the completeness and comparability of case findings and diagnostic and record keeping systems within and between countries.

In the United Kingdom, the Department of Health and Social Security collates certain information, including diagnosis, on every case admitted to public psychiatric hospitals in England and Wales. The standard of diagnosis in these epidemiological inquiry as done by Spier et al. (243), and M.

**Biological or Cultural Milieu**

One side of the coin in psychopathology is that and other side of the coin is the effect of environment.

Studies on family history, alcoholism are advanced (157, 282). Refinement possible to reexamine biology of depression and dementia.

The assessment of the cult which substantial examination of psychopathology is possible, proved a powerful approach to environment.

The pioneering work of Hillishing the association between the great volume of related studies and the onset of various symptoms and groups (167, 274a)

Although this area of mental illness has become a cross-national research.

4. ATTEMPTS TO IMI

Robert L. Spitzer and Jeanne.

**Computerized Diagnosis**

The limitations of the clinician's manual particularly the low reliability have led many investigators to use computerized systems in the clinical observations it.

1. Both authors are also affiliated with California Univers
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY 643

chiatric hospitals in England and Wales. In view of the relatively uniform and high standard of diagnosis in these hospitals, these data have proved very useful for epidemiological inquiry as exemplified by the age distribution studies of Hare et al (89), Spicer et al (243), and McDonald (149).

Biological or Cultural Milieu

One side of the coin in psychiatric epidemiological research is the assessment of psychopathology so that its distribution in time and space may be plotted. The other side of the coin is the assessment of factors that might be associated with this distribution.

Studies on family history, on genetic markers, and on associated disorders (e.g. alcoholism) have advanced knowledge on the etiology of manic-depressive disorders (157, 282). Refinement of indices for biological aging (14, 139) is making it possible to reexamine biological concomitants of age variation in the distribution of depression and dementia.

The assessment of the cultural milieu has reached a stage of sophistication at which substantive examination of the interaction between the ecological niche and psychopathology is possible. The concept of life events and their change has proved a powerful approach to isolating one of the pathogenic influences in the environment.

The pioneering work of Holmes, Wolf, Rahe, and their co-workers in establishing the association between stress associated with life events and illness, and the great volume of related work that followed, has led in the last few years to intensive study of the types and timing of life changes that are particularly likely to precede the onset of various psychiatric disorders (16, 46, 180) and to community surveys aimed at examining the relationship between the distribution of measures of psychiatric symptoms or disorders and of life events across various demographic groups (167, 274a).

Although this area of measurement is replete with methodological pitfalls (17), consistency in social consensus on the degree of change in given life events across cultures (191) makes a measure of life events a possible tool for cross-cultural and cross-national research.

4. ATTEMPTS TO IMPROVE PSYCHIATRIC DIAGNOSIS

Robert L. Spitzer and Jean Endicott

Computerized Diagnosis

The limitations of the clinical method for arriving at a psychiatric diagnosis, particularly the low reliability generally associated with routine clinical diagnosis, have led many investigators to explore the potential of computers for integrating the clinical observations into standard psychiatric diagnoses. Although some

3 Both authors are also affiliated with the Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York City.
efforts have been made to develop programs in which the patient interacts directly with the computer via a console, the inherent limitations of this approach have led most investigators in this area to employ models in which the basic observations of signs and symptoms are still made by clinicians. These are the clinical observations which are then introduced into a computer for summarization into a diagnosis.

There are several advantages to this kind of computer-generated diagnosis. First of all, there is the value of the computer’s necessarily perfect reliability in the sense that, given the same data, the computer program will always yield the same diagnosis. Secondly, the computer program can utilize rules developed from a larger and more diverse sample of actual patients than any single clinician can command. In addition, the rules by which a computer assigns a diagnosis are explicit and public. Finally, empirically based rules constitute at least potential advances in our scientific understanding of the complex relationship between symptom characteristics and diagnosis.

The most frequently employed models for computer diagnosis are either discriminant function or Bayes classification (11, 57, 171, 192, 240). Both of these models are empirically derived from data on patients for whom there are available clinical diagnoses. The data is usually psychiatric symptomatology, but may include demographic and historical information. These two statistical methods are used to develop optimal classification rules for the particular sample to which they were originally applied, the “developmental” sample. The derived rules are then applied to classify new cases. The major limitation of these models is the requirement of a very large sample size. Even with sample sizes in the thousands, which are rarely available, a very limited number of categories can be employed. A second limitation is the lack of generalizability of the rules to samples which differ greatly from the developmental sample.

The Bayes method has been applied to psychiatric classification by Birnbaum & Maxwell (11), Overall & Gorham (171), and Smith (240). The discriminant function method has been used by Rao & Slater (193), Melrose, Strobebl & Glueck (152), and Stetten, Altman & Uletti (237). Stetten and his group have applied this method to five hospitals in the Missouri automated record keeping system and developed a method for classifying patients into one of 12 diagnostic groups. The overall agreement between computer classification and clinical diagnosis was sufficiently high for them to make the system operational in the Missouri Standard System of Psychiatry used in Missouri state hospitals.

An alternative to the statistical models is the logical decision tree approach developed by Spitzer & Endicott (245, 246, 251). In this approach, which emulates the clinical differential diagnostic procedure, the computer program consists of a sequence of questions. The answer to each question is either “true” or “false” and rules out one or more diagnoses and determines which question is to be examined next. Some questions may ascertain the presence of a single critical sign or symptom, others whether a numerical score is in a certain range, and yet others whether a complex pattern of both signs and symptoms is present. The logical decision tree approach has the advantage of being independent of any specific body of data although, of course, not of accumulated clinical and research expe-
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY

patient interacts directly with this approach have led to the basic observations of this approach, and the clinical observation of the patient's reaction into a diagnosis.

The derived diagnostic rules, not perfect reliability in the individual case, will always yield the same result, but it is unlikely that any single clinician can consistently and reliably assign a diagnosis to a patient that is identical to that assigned to the same person by another clinician. This limitation was a result of at least potential diagnostic error and the lack of a common agreement on the definition of diagnosis.

Diagnosis are either described or named (92, 240). Both of these methods are available in clinical practice, but may involve different criteria. Statistical methods are based on large samples of data, and the derived rules are then applied to new cases. The requirement of a large sample, which is necessary in the thousands, which can be employed, is the requirement for the samples which differ significantly.

The discriminant function of Birnbaum and Stroebel (93, 244) has been shown to be useful in the discrimination of clinical groups. The method is based on the number of variables which are used in the diagnostic algorithm and the standard nomenclature of the Missouri Standard Diagnostic Criteria.

The decision tree approach to psychiatric diagnosis, which emulates the algorithmic process of a diagnostician, has been shown to be a useful tool in the diagnosis of psychiatric disorders. The decision tree approach is based on a series of questions, each of which is answered with a "true" or "false" answer. The process is to be examined by the clinician until a single critical sign or symptom is found, and yet others must be present. The logical structure is independent of any specific medical and research experience. For this reason the method can be used with different psychiatric populations, since it is not designed to be optimal for a particular population of psychiatric patients.

The two statistical models and the logic decision tree model have been compared on the same database (64). The logic decision tree model performed as well as the two statistical models (in terms of agreement with clinical diagnosis) on a large cross-validation sample which was similar to the developmental sample. The logic decision tree method performed better than the two statistical approaches on a sample drawn from a totally new population.

The logic decision tree approach has been employed in three programs of increasing complexity: DIAGNOS, DIAGNOS II, and DIAGNOS III by Spitzer & Endicott. The validity issue has been studied primarily by examining the agreement between the computer diagnoses and the diagnoses of expert clinicians, since the purpose of the program was to simulate expert clinical practice. When the computer and the clinician were limited to the same information (from research protocols of psychiatric ratings), the agreement between the computer and clinical diagnosis was as good as that between the diagnoses of pairs of clinicians (245, 246). However, in a study in which the clinicians had direct access to the patient or a narrative case record, the agreement between the clinicians exceeded agreement between the computer and the clinicians. The output of these programs has been judged to have sufficient validity for use in various research projects for describing samples of subjects, selecting subjects for experiments, in epidemiological and cross-cultural studies, and investigating problems in classification.

Dissatisfaction with the treatment relevance or standard diagnoses has led some investigators to employ computers to develop treatment-relevant typologies rather than to simulate the practices of expert clinicians using the standard nomenclature (93, 119, 172, 238). Each of these programs has been shown to have some value in predicting which patient will respond best to a limited number of drugs in samples similar to those used to develop the program. The general utility of such programs across samples of patients who differ from the samples from which the programs were derived has yet to be demonstrated.

In so far as the computer is used to simulate expert clinical practice in making a standard psychiatric diagnosis, the major constraints on further developments in computerized diagnosis appear to lie in limitations in the standard nomenclature itself. Although improvements in the raw data and the computer algorithms may lead to some improvement in computer diagnosis using the current ambiguous criteria, a quantum jump in the validity of computer diagnosis is dependent upon improvements in the system itself.

Specified Criteria for Psychiatric Diagnosis

In the usual diagnostic manual, such as the American Psychiatric Association's Diagnostic and Statistical Manual (DSM-II), diagnostic stereotypes are described under each category. The task of the diagnostician then becomes that of selecting the diagnostic category in which the stereotype most closely resembles the characteristics of the patient being diagnosed. In addition, the standard nomenclature
includes many categories for which there is little validity evidence in terms of discriminating from other conditions. Long-term follow-up studies, specific treatment response, typical course, and elevated familial incidence.

The Renard hospital group in St. Louis (54) has developed a diagnostic system which limits itself to those conditions for which it believes there is good research evidence for the validity of a category.

Spitzer, Endicott, collaborative study at Beth Israel, and criteria in an effort to categorize patients with major depressive disorder (254). The study has been included in the Diagnostic and Statistical Manual of Mental Disorders (4th ed.), published by the American Psychiatric Association.

Table 1 Specified criteria for major depressive illness (A through E required)

<table>
<thead>
<tr>
<th>A. Dysphoric mood characterized by symptoms such as the following: depressed, sad, blue, hopeless, down in the dumps, empty, “don’t care,” irritable, fearful, worried. The dysphoric mood must be prominent and relatively persistent but not necessarily the most dominant symptom.</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. At least 5 of the following symptoms for definite and 4 for probable (for past episodes because of memory difficulty, the criteria are 4 and 3 symptoms):</td>
</tr>
<tr>
<td>1. Poor appetite or weight loss (2 lbs. a week or 10 lbs. or more a year when not dieting) or increased appetite or weight gain.</td>
</tr>
<tr>
<td>2. Sleep difficulty or sleeping too much.</td>
</tr>
<tr>
<td>3. Loss of energy, fatigueability, or tiredness.</td>
</tr>
<tr>
<td>4. Psychomotor agitation or retardation (but not mere subjective feeling of restlessness or being slowed down).</td>
</tr>
<tr>
<td>5. Loss of interest in usual activities, or decrease in sexual drive, (do not include if limited to a period when delusional or hallucinating).</td>
</tr>
<tr>
<td>6. Feelings of self-reproach or excessive or inappropriate guilt (either may be delusional).</td>
</tr>
<tr>
<td>7. Complaints of or evidence of diminished ability to think or concentrate, such as slow thinking or mixed-up thoughts (do not include if associated with obvious thought disorder).</td>
</tr>
<tr>
<td>8. Recurrent thoughts of death or suicide, including thoughts of wishing to be dead.</td>
</tr>
<tr>
<td>C. Depressive features of illness lasting at least 2 weeks.</td>
</tr>
<tr>
<td>D. Sought help from someone during the dysphoric period or had impaired functioning socially with family, at home, or at work.</td>
</tr>
<tr>
<td>E. None of the following which suggests schizophrenia is present:</td>
</tr>
<tr>
<td>1. Delusions of control or thought broadcasting of any duration if definitely present.</td>
</tr>
<tr>
<td>2. Hallucinations throughout the day for several days, or intermittently throughout a 1-month period unless content is clearly related to lowered or elevated self-esteem.</td>
</tr>
<tr>
<td>3. Auditory hallucinations in which either a voice keeps up a running commentary on the patient’s behavior as it occurs, or 2 or more voices converse with each other (of any duration as long as definitely present).</td>
</tr>
<tr>
<td>4. At some time during the period of illness had delusions or hallucinations for more than 1 month in the absence of prominent affective (manic or depressive) symptoms (although typical depressive delusions, such as delusions of guilt, sin, poverty, nihilism, or self-deprecation or hallucinations of similar content are permitted).</td>
</tr>
<tr>
<td>5. Preoccupation with a delusion or hallucination to the relative exclusion of other symptoms or concerns (other than delusions of guilt, sin, poverty, nihilism, or self-deprecation or hallucinations with similar content).</td>
</tr>
<tr>
<td>6. Definite instances of schizophrenic thought disorder.</td>
</tr>
</tbody>
</table>

Table 2 Kappa coe
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY  

Evidence for the validity of the conditions as diagnostic entities. The system provides for explicit criteria which are in the form of sets of items, a specified number of which are necessary or sufficient to make a diagnosis. A consequence of this approach is the recognition of the need for an "undiagnosed psychiatric disorder" category for the patients who do not meet any of the specified disorders, yet exhibit significant psychopathology.

Spitzer, Endicott, and Robins (personal communication, 1974), as part of a collaborative study of the psychobiology of depression, have expanded the Renard system to include other categories and have modified and refined the specified criteria in an effort to improve reliability. An example of the criteria for one category, Major Depressive Illness, is shown in Table 1.

The reliability of these diagnostic criteria has been evaluated based on both interviews of patients and case records. The reliabilities obtained with 104 newly admitted psychiatric inpatients from four participating facilities were generally considerably higher than those reported in other diagnostic reliability studies (254). The study of case records used 120 patients from the New York State Psychiatric Institute and permitted a comparison between the reliabilities obtained when research assistants used the specified criteria and when a different set of experienced clinicians used the official nomenclature of the American Psychiatric Association. The results of the case record study are shown in Table 2.

Table 2  Kappa coefficients of agreement for major diagnosis on 120 case records

<table>
<thead>
<tr>
<th>Diagnostic category</th>
<th>Kappa using specified criteria</th>
<th>Kappa using DSM-II criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major groupings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective</td>
<td>.70</td>
<td>.48</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>.84</td>
<td>.48</td>
</tr>
<tr>
<td>Other</td>
<td>.74</td>
<td>.61</td>
</tr>
<tr>
<td>Specific categories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonaffective schizophrenia</td>
<td>.70</td>
<td>.49</td>
</tr>
<tr>
<td>Schizoaffective schizophrenia</td>
<td>.24</td>
<td>.37</td>
</tr>
<tr>
<td>Manic illness</td>
<td>.78</td>
<td>.37</td>
</tr>
<tr>
<td>Major depressive illness</td>
<td>.66</td>
<td>.25</td>
</tr>
<tr>
<td>Minor depressive illness</td>
<td>.56</td>
<td>.07</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>.66</td>
<td>.06</td>
</tr>
<tr>
<td>Drug dependence</td>
<td>1.00</td>
<td>.32</td>
</tr>
<tr>
<td>Obsessive-compulsive</td>
<td>.65</td>
<td>.42</td>
</tr>
<tr>
<td>Anxiety neurosis</td>
<td>.52</td>
<td>.32</td>
</tr>
<tr>
<td>Phobic neurosis</td>
<td>.66</td>
<td>.20</td>
</tr>
<tr>
<td>Borderline psychosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undiagnosed</td>
<td>.59</td>
<td></td>
</tr>
</tbody>
</table>

a Considered equivalent to any psychotic affective illness in DSM-II.

b Considered equivalent to Neurotic Depression in DSM-II.

c Neither of the raters used this category.
using a measure of agreement called kappa, which indexes agreement for nominal categories corrected for chance agreement (244). Not only are virtually all of the reliabilities higher for the research assistants using the specified criteria, but the values obtained are higher than those ever reported in diagnostic studies of heterogeneous samples of patients.

To further increase the reliability of this set of diagnoses, an instrument has been developed by Spitzer and Endicott, called the Schedule for Affective Disorders and Schizophrenia (SADS). The SADS has a structured interview and a set of rating scales to enable a clinician to collect the information necessary to make the diagnostic distinctions provided in the list of diagnostic criteria. In addition, a computer program will be developed which will use the specified diagnostic decision rules to classify patients on the basis of the individual items. It is quite clear that one way of increasing reliability of diagnoses is to provide more specific criteria for each diagnostic category.

5. A PSYCHOPHYSIOLOGICAL APPROACH TO PSYCHOPATHOLOGY

Samuel Sutton

The rationale for applying an experimental approach to mental disorders derives in part from the need for objectifying and measuring the observations of the clinician on the characteristics and behavior of patients and, more deeply, from the hope of finding some kind of indicator of difference in central nervous system functioning that might define a group of patients. The concern of this section is primarily with the latter aim. In remarking some time ago on the subtlety and complexity of the etiological factors in mental illness, as well as on the kinds of artifact which may arise in biochemical research involving the analysis of body fluids and tissues, Kety (112) noted that it would be highly desirable if some clues as to what should be looked for could be supplied to the biochemist from other fields—not least of all, psychology.

It might seem paradoxical that one would turn to psychological rather than physiological responses when one's primary goal is to obtain some reflection of central nervous system functioning. However, to the extent that the physiological variables are sensitive to the psychological state of the subject (263), no advance is made merely by moving from the psychological to the physiological level. Thus it is true that schizophrenic patients may be shown to have unusual evoked potential waveforms to stimuli whose identity they did not know in advance (133). However, from the point of view of providing etiological clues, such findings may be viewed as redundant support of available psychological data, e.g. reaction time, which have shown that the response of schizophrenic patients to the condition of uncertainty is different from that of normal subjects (266). Both the physiological and psychological data share a common problem of validity. In neither case do we know whether the difference between groups is in the way they respond to uncertainty, or whether other differences between groups, such as the level of attention or motivation solved at the psychosocially focused. Only then performance be illus.

The fundamental that patient groups: the the sources of the patient has a deficit psychomotor function to do with these fulfill may be less coopered to may be preoccupy apparent completely meaning patient, the complete may obtain erroneous and too concerned to high degree of an problem in obtaining different from the f been most often rai.

Several strategies and obtaining more will now be present

Simplicity of the

Simple reaction time stimuli (242, 257) when one of sever procedure. Nevert stimuli inescapably appreciating the difference the same stimuli pr be too difficult or to the usual uninterpr.

However, the re most hospitalized p ulti. This brings intc

Own-Control Des

The rationale of the or trivial factors ent would presumably difference scores or also lead to serious
scees agreement for nominal
test, only are virtually all of the
items specified criteria, but the
of diagnostic studies of
uses an instrument has
Schedule for Affective
Structured interview and a
information necessary to
of diagnostic criteria. In
which will use the specified
of the individual items. It
diagnoses is to provide more

mental disorders derives
the observations of the
and, more deeply, from
central nervous system
concern of this section is
go on the subtlety and
to the kinds of
the analysis of body
desirable if some clues
biochemist from other

psychological rather
contain some reflection of
that the physiological
ject (263), no advance is
physiological level. Thus it
usual evoked potential
tance (133). However,
findings may be viewed
reaction time, which
of the condition of uncertainty
physiological and
neither case do we
they respond to un
such as the level of

attention or motivation, determine the findings. The problem of validity can be
solved at the psychological level by eliminating or controlling the effect of extraneous psychological processes on the behavior on which the experimenter is
focused. Only then might the obtaining of physiological correlates of deviant
performance be illuminating with respect to etiological questions.

The fundamental difficulty in the interpretation of findings arises from the fact
that patient groups almost universally perform more poorly than normals. But it is
the sources of the poor performance that are unclear. It may in fact be true that the
patient has a deficit in information processing, or in perceptual sensitivity, or in
psychomotor function. However, the poor performance may have nothing directly
to do with these functions and may arise for quite different reasons. The patient
may be less cooperative or less attentive. He may be concerned with his troubles,
may be preoccupied with his hallucinations, and may care very little about this
apparently meaningless task we put before him. Given the disturbed state of the
patient, the complexities of the task may be too much for him. Paradoxically we
may obtain erroneous results because the patient is too motivated, too anxious,
and too concerned with the results of his performance, and as is well known, a very
high degree of anxiety and motivation often lead to poor performance. The
problem in obtaining valid data in these situations is in a sense not logically
different from the problems of controlling for “faking” or malingering which has
been most often raised in relation to responses to questionnaires (228).

Several strategies have been developed for coping with some of these problems
and obtaining more valid indicators of central nervous system differences. These
will now be presented (98a, 264).

Simplicity of the Task

Simple reaction time can be used to reflect fairly subtle discriminations between
stimuli (242, 257). The subject's task is only to lift his finger as rapidly as possible
when one of several stimuli is presented. No discrimination is required in
this procedure. Nevertheless, a finding of differences in reaction time to different
stimuli inescapably leads to the inference that the nervous system is capable of
appreciating the differences among them. The advantage in such a strategy is that
the same stimuli presented as a discrimination may for whatever reason prove to
be too difficult or too subtle a task for certain groups of patients. This would lead to
the usual interpretable "poorer" performance.

However, the reaction time of schizophrenic patients, and to some extent of
most hospitalized psychotic patients, is generally slower than normals to all stim-
uli. This brings into play a second strategy.

Own-Control Design

The rationale of the own-control approach is based on the hope that confounding
or trivial factors enter equally into the two or more measures being compared. This
would presumably permit the ruling out of confounding factors by the use of
difference scores or some other maneuver. While this can be a powerful tool, it can
also lead to serious misinterpretations. For example, if the measures being used

involve tasks of different degrees of difficulty, then it is possible that the level of difficulty can interact with differences between groups in motivation or attention. Performance on the difficult task may be more severely handicapped by low motivation than performance on the easy task; thus a difference between groups in motivational level would make performance on the easy task an inappropriate correction factor for comparison of groups on the more difficult task. Furthermore, this process of correcting for performance at one level by performance at another level has a number of other pitfalls, and one cannot with impunity take difference scores or use even covariance analysis indiscriminately in all cases (66, 264).

Patients and normals must be tested on identical experimental conditions to avoid bringing range effects into play (187). Furthermore, care must be used to randomize experimental conditions, by trial if possible, so that subjects cannot assume different attitudes to various experimental conditions, e.g. the attitude that “this is a challenging condition I will try harder”; or “this is a very difficult condition, I give up on this one”; or “this is an easy condition, I don’t have to try so hard”; and so on. Even with all these provisos, a series of sequential experiments may be necessary to establish the interpretation of a particular finding (188). For example, Sutton & Zubin (266) reported that schizophrenic patients have longer reaction time than normals when the immediately preceding trial is in a different sensory modality than when the immediately preceding trial is in the same sensory modality, even though the stimulus of the previous trial occurs as long as several seconds earlier. This appears to imply a difference in bias or expectancy or segmental set on the part of the patients. However, in a subsequent experiment it was shown that the greater crossmodal retardation continues to be found in such patients even when the subject is told prior to each trial what the sensory modality of the stimulus will be (273a).

Controlling the Criterion Variable

Signal detection theorists (79) have been able to demonstrate that they can obtain two independent measures in threshold situations: one is a measure of sensitivity called \( d' \); and the other is a measure of criterion \( \beta \). The basic advance consists of the fact that it can be demonstrated both logically and experimentally that these two measures can be obtained independently of each other. This is an important step forward since many patients tend to be more cautious than normals. It is indeed this cautiousness which accounts for the earlier reports that schizophrenic patients were less sensitive in their ability to detect flicker. What Clark, Brown & Rutschmann (28) have shown is that, rather than being less sensitive, the patients appear to have a need for seeing definite flicker before they are willing to say that the light is flickering. With classical psychophysical methods such a bias yields an apparently poorer sensitivity. When measures of sensitivity are obtained while between schizophrenia and darkness may be obtained.

Clark and his colleagues used classical psychophysical methods to the criterion which required further study.

Several other methods yielded results found no difference in obsessive and normal which permitted making a decision than four times stronger at the method of thermal sensitivity. The psychiatric patients needed this technique.

The psychiatric patients differed in their use of sensory sensitivity. Differences in the use of stimulus presented while normals in a half of electric shock under these conditions of paired samples not under more threat. Neuroticism score found signal detection which involving on manifest differences in psychophysical differences were significant. Despite the finding that criterion therefore, we should negative results were obtained.

In psychiatric patients more often that

---

4 Beta is the term more commonly used to refer to the signal detection measure of criterion. Actually, what is usually computed from hits and false alarm rates is \( L_x \). For the difference between \( L_x \) and \( \beta \), see Swets, Tanner & Birdsall (267).
are obtained which are independent of the criterion variable, there is no difference between schizophrenic patients and normals in sensitivity to the rate at which light and darkness must alternate before flicker is detected. The pioneering papers by Clark and his colleagues (27, 28, 31) have cut the underpinnings from a great deal of research comparing the sensitivity of patients and normals which was done by classical psychophysical techniques. If the difference in performance is in fact due to the criterion variable, then it is the criterion variable and not sensitivity which requires further study.

Several other investigators have since reported that the use of signal detection methods yields a finding of no differences between groups. Milner et al (158) found no differences in either $\alpha$ or $\beta$ for the detection of auditory tones in noise by obsessional and nonobsessional patients. However, in one experimental condition which permitted the subjects to ask for repeat observations of the signal before making a decision, it was found that obsessional patients asked for repeats more than five times as often. Clark & Rubin (30) found that while the constant stimulus method showed that the psychiatric patients (who were older) had poorer thermal sensitivity than the normals, a signal detection theory analysis of the same data showed no differences in $\alpha$, but rather differences in the criterion variable. The psychiatric patients were more conservative.

Not all use of signal detection methods has led to negative results with respect to sensitivity differences between psychiatric or retarded groups and normals. Kozh & Vaytkavichyus (122) found differences between schizophrenic patients and normals in alpha suppression and electromyographic responses to auditory stimuli presented in a signal detection task. Kuechler & Dodwell (128) found that while normals increased their $\alpha$ for the detection of a tone in noise under the stress of electric shock, neurotic and schizophrenic patients showed a decrement in $\alpha$ under these conditions. Boissonneault et al (13) found $\alpha$ differences in the resolution of paired flashes only under a high exercise-induced activation level—but not under moderate or low activation levels—between extroverts with a high neuroticism score and introverts with a low neuroticism score. Subkis (262) found signal detection theory of use in the analysis of results of published experiments which investigated classical discrimination conditioning in relation to ratings on manifest anxiety.

In a comparison of retarded and normal adults, Bernstein & Day (10) found differences between groups in size constancy with both classical and signal detection psychophysical methods. In the Price & Eriksen (189) study of size constancy, differences were found between nonparanoid schizophrenic patients, paranoid schizophrenic patients, and normal controls with the signal measure, despite the findings of no differences with the classical psychophysical measures. When criterion is uncontrolled, no real inference about sensitivity can be made; therefore, we should not be surprised by whatever permutation of positive and negative results we obtain with classical psychophysical and detection theory methods.

In psychiatric illness, differences between groups in the criterion variable are more often than not viewed as an artifact which interferes with the evaluation of
differences in sensitivity. Signal detection methods, however, have also been applied where the primary goal of the study was to look for criterion differences between groups, rather than for sensitivity differences. In one study, Carney (21) found no differences between delusional and nondelusional schizophrenic patients in their response to the use of monetary payoff to manipulate the criterion level. The assessment of the effects of aging presents a contrast to the assessment of psychiatric illness. While investigators of aging are also concerned with the evaluation of sensitivity, the hypothesis that criterion differences such as conservativeness or cautiousness are associated with aging has been of particular interest to investigators. Craik (42) reviewed the evidence from his own earlier study as well as the available literature which used signal detection methods. While support was found for high βs (greater cautiousness) to be positively associated with age, there were also negative and contradictory findings. Evidently the nature of the task enters as a factor in determining whether or not a relationship with age will be found as well as the direction of the relationship (29, 30, 78, 203).

The evaluation of the effects of drugs on performance in both normal subjects and patient groups raises many of the same difficulties that comparison of normal and patient groups raises. Here, too, the problem which arises is whether the drug has affected the subject’s sensitivity, his criterion, his attention level, his motivation, etc. An increasing number of drug studies using signal detection theory methods in an attempt to disentangle some of these effects are beginning to appear (48, 49, 82, 143, 164, 194–197, 225, 290).

Problems of comparing patient groups with normals are by no means all solved by the use of signal detection methods. Thus Rappaport et al. (196), who used signal detection methods, discounted a finding of poorer sensitivity in nonparanoid schizophrenic patients than normals when they found that at still lower stimulus intensities where the discrimination is more difficult, the same patients were no less sensitive than the normals. They argue that such contradictory findings reflect the fact that signal detection methods cannot cope with the variable attention level characteristic of certain patient groups. However, strictly speaking, they did not use a signal detection format since a “no” response was defined as a failure to make a response (48).

Accuracy Indicators

While the subject’s response in the “yes-no” or one-interval signal detection format is a forced choice in that the subject must make a decision, other forced-choice designs have different properties. Thus in the three-interval forced choice, the subject must in each trial decide whether the stimulus was in the first, second, or third observation interval. Such a procedure differs from the one-interval detection theory procedure in two important ways. First, it provides the subject with a comparison in each trial—namely, the observation interval in which the stimulus was present compared with the two observation intervals in which the stimulus was not present. Second, the subject is in effect being forced to respond “present” following each trial, eliminating the usual avenue for expressing cautiousness or laxness criterion free; criterion.

But it should be noted that the stimuli are equated in relation to the stimulus and the response bias. In a short-term memory study examining the hit ratio, it is utilized the end point of the normal range. Han and Cohen (116) found that detecting electric stimuli Boren & Nakai and chronic nonspecific traction. They found that the distraction the sensitivity of the acute patients and poor whether a tachistoscopic signal letter was performed more frequently in patients with a nonchislo-geometrical figures noise. All forced-choice thresholds obtained, and the schizophrenic patient matched with nonschizophrenic study of short-term memory.

Walker & Bircz between ages 7–11 matched an auditory age made fewer correct responses.
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY 653

However, have also been found for criterion differences in the one-interval forced-choice procedure. In one study, Carney (21) found that schizophrenic patients could manipulate the criterion to the extent that they could judge the correct stimulus even to the assessment of the observer. The present study was concerned with the evaluation of three observers, such as conservatorship cases of particular interest to the author and the particular criteria in a previous study as well as other factors. While support was found for the idea that normal subjects were not as upset as normal subjects in the second study, the nature of the task of the psychopath not the age of the psychopath might be expected to affect the results (41, 203).

In a study comparing normal and schizophrenic patients, it was found that when the drug did not have a positive influence on the patients, the high detection level, his motivation. The signal detection theory was used to explain these findings.

The data on the patients' instructions were also analyzed. Although no means all patients were able to understand the instructions, the same patients were not always able to understand the instructions at lower levels of signal stimulus. However, strictly speaking, the decision to use a "no" response was a matter of judgment.

The three-interval signal detection procedure was used to determine whether chronic schizophrenia patients could manipulate the criterion to the extent that they could judge the correct stimulus. The three observers were not significantly different from the one-interval procedure. In the three-interval procedure, the subject was forced to respond in a different way for expressing caution or laxness of criterion. Thus the three-interval forced-choice procedure is criterion free; criterion in the usual sense does not enter nor can it be measured. But it should be noted that it is not free of response bias. The subject may not distribute his responses equally among the three observation intervals (although the stimuli were equally distributed). When this occurs, it can be corrected for by examining the hit rate separately for each interval, thus correcting for the subject's response bias. In a spatial forced choice, Calfee (10) showed that subjects did not utilize the end positions and the center positions in the same way. This was a short-term memory task and subjects remembered stimuli in the end positions better than stimuli in the center positions. As a result, when they were uncertain they more often picked center positions. This bias was the same for retarded as for normal children.

As with signal detection methods, few investigators are using forced-choice methods to compare groups. Emmerich & Levine (46) found that schizophrenic patients had less sensitive auditory thresholds than normals. Bruder et al (47) found that only patients with affective psychoses had less sensitive auditory thresholds than normals. The sensitivity of schizophrenic patients was found to be within the normal range. Hare (91) found that psychopathic criminals were less sensitive in detecting electric shock than nonpsychopathic criminals.

Broen & Nakamura (15) compared the auditory sensitivity of acute paranoid and chronic nonparanoid schizophrenic patients under distraction and no distraction. They found no difference in the no-distraction condition, but under distraction the sensitivity of the chronic nonparanoid patients deteriorated while that of the acute paranoid patients did not. Neale et al (169) found that there were no differences in the ability of normals, good premorbid paranoid schizophrenic patients, and poor premorbid nonparanoid schizophrenic patients to identify whether a tachistoscopically displayed letter was a T or an F. However, when the signal letters were imbedded in an 8 letter matrix, both schizophrenic groups performed more poorly than the normals. Stilson & Kopell (259) compared normals, nonschizophrenic, and schizophrenic patients in their ability to identify geometrical figures against a background of visual noise as opposed to no visual noise. All forced-choice thresholds were elevated by visual noise; the schizophrenic thresholds were most elevated, the nonschizophrenic patients were intermediate, and the normals were least elevated. Smith (239) reported that chronic schizophrenic patients showed an impairment in short-term memory when compared with nonchronic schizophrenic patients. Also see Shallice & Warrington (229) for a study of short-term memory in a case of conduction aphasia.

Walker & Birch (276) had schizophrenic and nonschizophrenic children between ages 7-11 select from three visual dot patterns the one which best matched an auditorily presented temporal pattern. Schizophrenic children at each age made fewer correct matches.

Better Patient Performance

While the use of forced-choice methodology controls the criterion problem very well and also gives us a safeguard against falsely attributing to patients a better performance than they can actually accomplish, it does not solve the fundamental problem. One can argue very persuasively that the patients who have poorer thresholds than normals were simply less attentive or less motivated than normals. Presumably, according to this argument, had the patients been in a state which permitted them to be fully attentive and cooperative they might have obtained just as good thresholds as the normals. There is only one satisfactory answer to this argument. If we could demonstrate in a particular task that patients were able to perform better than normals, it would then not be possible to attribute the better performance for that task to poorer attention and cooperation on the part of the patients. There are so far only a few instances of reported better performance on the part of patients.

Bruder et al (see footnote) presented randomly either a single click which was 25 dB above each subject's absolute threshold, or the 25 dB click followed 15 msec later by a click which was 10 dB above the subject's absolute threshold. The subject lifted his finger from a key as rapidly as possible in response to the presentation of either stimulus package. They found that reaction time of normal subjects to the paired clicks was only very slightly faster than reaction time to the 25 dB click by itself. However, for patients with affective psychoses, reaction time to the paired clicks was much faster than reaction time to the 25 dB click by itself. Here then is evidence that these patients were benefiting more from the presence of the 10 dB click than were the normals. The patients with affective psychoses were more "sensitive" to the presence of the 10 dB click and in that sense were performing better.

Collins (33) presented an even more convincing case of better performance. There were two 2 msec flashes of light, which in one stimulus package were separated by 2 msec giving a total duration of 6 msec, and in another stimulus package were separated by 0 msec—that is, a continuous 4 msec flash of light. The two stimulus packages had the same total amount of light energy; however, the energy was distributed differently over time. As is known from Bloch's Law, if the total stimulus duration of both packages is shorter than a value known as critical duration, then the level of response to these two packages of light will be identical. If, however, one of the stimulus packages exceeds critical duration for the class of response under consideration, then the level of response to the two equal-energy packages will be different. These two stimulus packages were presented in random sequence, and subjects lifted their finger as rapidly as possible whichever stimulus package appeared. For normals and nonschizophrenic psychiatric patients no reaction time differences were found between the two stimulus packages. However, schizophrenic patients showed significantly longer reaction times to the 6 msec package of light than to the 4 msec package of light. From this it is inferred that for schizophrenics, critical duration is shorter than 6 msec. Presumably, for this group some process contains less effect...
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY 655

determines the criterion problem very well, indicating that patients' abilities are more similar than different. The criterion problem is well established for patients who have lower or less motivated than normals.

patients been in a state which might have obtained just one satisfactory answer to this task that patients were able to attribute the better cooperation on the part of the patient group.

either a single click which was the 25 dB click followed 15 msec absolute threshold. The subject's response to the presentation of a single 25 dB click by itself, reaction time to the paired 10 dB click by itself. Here then is the presence of the 10 dB effective psychotic were more than in that sense were performing in a state of better performance.

stimulus package were separated in another stimulus package a flash of light. The two stimuli were equal in energy; however, the energy was from Bloch's Law, if the total energy in a value known as critical power is the same for the two equal-energy stimuli. There were presented in random order. It is possible that the use of a stimulus in one is not possible that the use of two stimulus packages. However, reaction times to the 6 msec light. From this it is inferred that 6 msec. Presumably, for

this group some portion of the 6 msec package is not fully integrated. It therefore contains less effective energy and reaction time is longer.

These experimenters (33, 264) make no claim that a shorter critical duration is biologically “better” than a longer critical duration—in fact, the reverse may well be true. But whichever is functionally more advantageous, the schizophrenics in this design make a discrimination, in the sense of different reaction time to the two packages, which normals are unable to make. (The two stimulus packages are also perceptually identical for normals; comparable perceptual studies are not available for patients.) It would be difficult to attribute such a group difference to defective attention or cooperation on the part of the patient group.

Two studies of behavior at a much more complex level report findings which may also be interpretable as better performance by patients. Goldfarb & Braundstein (75) reported that the speech of schizophrenic children is less disturbed than the speech of normal children by a delay in auditory feedback while speaking. Levine et al (132) reported a similar finding for schizophrenic adults. However, an earlier study in schizophrenic adults by Sutton et al (265) reported negative results.

Iterative Procedures

All the above strategies are addressed to the design of the tasks that the subjects perform. However, there is at least as serious a problem in defining the independent variable, namely the group to which each subject belongs. For the definition of patient groups, there is the problem of the lack of objectivity of the criteria and the lack of agreement among clinicians. Less well recognized is the fact that a “normal” population may often contain a significant proportion which is far from normal. For the classification of both patients and normals, the use of standardized rating scales which have good reliability is a distinct advantage. Kriege (126) reported that in her samples only half of the hospital-diagnosed schizophrenics were diagnosed schizophrenic by rating scale procedures. Even more surprising is the fact that only half her normals, recruited through the usual channels, were diagnosed as normal by the rating scale procedures. When her group comparisons were confined to the “pure” schizophrenics and the “pure” normals, not only were the reported differences in crossmodal retardation larger, but also intragroup variability was found to be smaller. Along similar lines, in the Collins study (33), it was those schizophrenics who rated high in “speech disorganization” who demonstrated shorter critical duration. By similar post hoc analyses, Satterfield (223) showed that it was depressives with a family history of mental illness who were most different from normals in their evoked potential waveforms.

Because such findings are post hoc discoveries, they must be subjected to replication before being accepted. Sutton (264) has argued that it is through such a series of sequential experiments that the experimental psychologist may make the

greatest contribution toward the "purification" of groups and thus to the problems of diagnosis and etiology. A series of sequential experiments permits the utilization of an "iterative" strategy. For example, after replication of the initial finding with respect to critical duration, the measure of critical duration can then be used as the independent variable in selecting subjects to see whether it is only schizophrenic "speech disorganization" that is associated with shorter critical duration. Other objective, physiological, and clinical techniques can then be brought to bear in the attempt to define and describe subpopulations of patients who by other procedures are inseparable from poorly defined larger categories.

ACKNOWLEDGMENTS

The author is indebted to Dr. W. Crawford Clark for advice and consultation. Preparation of the manuscript was in part supported by grant MH 18422 from the National Institute of Mental Health, United States Public Health Service, and by a project grant from the New York State Department of Mental Hygiene.

SUMMARY AND CONCLUSIONS

Joseph Zubin

With the tools described in the prior sections we can proceed to examine some of the scientific models that have been proposed for the etiology of the mental disorders.

Descriptive Model

In the descriptive area we have made considerable progress in demonstrating that systematic structured interviews of high reliability can be successfully applied to answering such questions as (a) whether more schizophrenic than affective disorders are admitted to the hospitals in London than in New York (see p. 637); (b) whether diagnoses arrived at by computers are comparable in reliability to those arrived at by clinicians (see section by Spitzer & Endicott); (c) whether patients who differ from their peers in the processing of information in the central nervous system show any characteristic difference in their psychopathology (see pp. 654–55); and (d) whether different treatments affect psychopathology differentially (95).

In addition to the quantitative approaches some notable progress has also been made in the historical documentation of qualitative description of psychopathological phenomena. Thus Sir Aubrey Lewis (134–136), one of the collaborators on the bilateral studies of diagnoses in the United States and the United Kingdom has traced the varying phenomenological description of the terms of anxiety, paranoia, and endogenous-exogenous historically and indicated their clinical significance in each period, including their present connotations. He has similarly described the development of the diagnostic schemas (137). Fleiss (62), on the other hand, with the help of his clinical colleagues has succeeded in developing a dimensional analysis of psychopathology based on the results of systematic inter-

views which yield: of affect separate before in the fact supported by fact already demonstr patients.

The diagnostic gation. Thus, in a patients whose in and applied a sec for the primary coefficients for s secondary diagnc agreement, to 911.

A second esti tion of diagnose whom they inter they interviewed could be regard of agreement in ability. It still re state hospital cli clinicians in Lon of schizophrenic diagnosis. Accp detection theory clinicians differo out to be very sm to apply the lal Kingdom—were English psychiat affective type, bi temic label and produced the o communciation.

Perhaps one interviews is the the world (289) systematic structur the grow centers. In a rep the 12 identified but not a single i point at a scor identified (only made positives)
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY
psychiatrists making these diagnoses, though indigenous to their own particular cultures, were all trained in the West and were viewing the patients through a Western-tinted screen. Nevertheless, the presence of similar groups of patients across this wide cultural spectrum would argue for the ubiquity of schizophrenia and presents a quandry for those who regard the diagnostic nomenclature as a myth. The items forming the 12 point scale included: 1. absence of waking early, depressed facies and elation, which represent affective components unfavorable to a diagnosis of schizophrenia; 2. presence of delusions (widespread, bizarre, nihilistic, and thoughts aloud); 3. disturbance of communication (incoherent speech, poor rapport, and unreliable information); and 4. poor insight and restricted affect.

Etiological Models

The etiological models that have been proposed for the mental disorders range from field-theory-based models, embedded in the ecological niche that the person occupies, to atomistic (biological) models of the genetic variety. Leaning towards the ecological model but containing also biological elements are two psychologically based models—the developmental model and learning theory model—while the two physiologically based models—the internal environment model and the brain function model—lean more in the direction of the atomistic approach.

This composite review has dealt with the learning theory model in the section by Salzinger and with the neurophysiological model in the section by Sutton. We shall not deal with these further, but instead will devote some attention to the ecological, developmental, genetic, and internal environment models.

The ecological model for the etiology of mental disorders assumes that their cause is to be sought in the stressors affecting the ecological niche which the person occupies. A full taxonomy of the parameters of the ecological niche and the stress they produce is not yet available, but such factors as low socioeconomic status have been implicated either as causes or as effects of mental disorder. A current investigation by the Dohrenwends (45) may throw light on this question. Assuming that old American stock who are now found in the low socioeconomic strata have gravitated there because of their lack of capacity to rise with the upward mobility thrust of the general population, they would tend to have higher rates of mental disorder than the newly arrived immigrants who contain a large proportion of capable people in their midst who have not yet had the opportunity to rise. On the other hand, if we assume that the stressful environment is the cause of mental disorder, we would expect the newly arrived immigrants who have risen to the higher levels to have more mental disorder than their peers of the old American stock who have not had to struggle to rise.

Perhaps the most striking evidence for the possible role of the ecological model comes from H. B. M. Murphy (166), who has found that whenever ethnic groups containing both Protestant and Catholic subgroups are compared by religious subdivisions for first admission rates for schizophrenia, the Catholic subgroup invariably has the higher rate. Whether this is attributable to Catholic upbringing or whether some artifact such as greater utilization of hospitals is the basis for this finding is not yet determined.

With regard to the entire field. An model, especially case for the degree unfortunately pro-family interaction deflated by Manfr their families (12). the Bürgholzli H patients from the bands and their re model is the lack o of preschizophren visual reasoning f factor achieved by in 1960, and who s

The genetic mc diagnostic categori (274). The introducic and environme the genetic comp victory of the gen schizophrenia, the spreading from gh schizoid and to t contributions of th reaction time exhig genetics. 7

Genetics has als Although the ear genetics in the tra volved were not kr on the X chromos bipolar depressiv the recent symposi

The internal env again in the foref environment mod flow and metabolic ering some unique synaptic mechanisms. Furthermore, 1 for inappropriate

7 J. L. Van Dyke, I and rearing on reacti
With regard to the developmental model, Garmerzy (72) has recently reviewed the entire field. Among the most vocal proponents of the role of the developmental model, especially with regard to family structure, is Lidz (138). He makes a good case for the degree of warping that may occur in families of the deviant type, but unfortunately provides no normal control groups for contrast. The dire role that family interaction plays in the development of schizophrenia has recently been deflated by Manfred Bleuler’s long-term study of 208 schizophrenic probands and their families (12). It is interesting to note that Manfred Bleuler, who grew up at the Burgholzli Hospital, found no help in diagnosis or in understanding his patients from the Rorschach technique which was administered to all the probands and their relatives. One of the more striking findings in the developmental model is the lack of intimacy which characterizes the adolescent friendship pattern of schizophrenics (125). Another set of findings are the poorer scores on a visual reasoning factor, on a social conformity factor, and on an impulse control factor achieved by a group of Project Talent subjects at the time they took the tests in 1960, and who subsequently were hospitalized for mental disorder (8).

The genetic model is spreading from the field of schizophrenia into other diagnostic categories such as depression (56), neurosis (231), and even criminality (274). The introduction of the study of adoptees for contrasting biological (genetic) and environmental (rearing) factors has demonstrated the strength with which the genetic component will make itself felt independently of rearing, but this victory of the geneticists has been gained at some sacrifice. The inheritance of schizophrenia, though undeniable, is no longer specific, but lies along a continuum spreading from genotypes whose phenotype is quite normal to those who are schizoid and to the more severely afflicted (114). One of the most interesting contributions of the introduction of the adoption studies is the finding that the slow reaction time exhibited by schizophrenics might be due to rearing rather than to genetics.7

Genetics has also demonstrated its importance in the field of affective disorders. Although the earlier twin and family studies had stressed the importance of genetics in the transmission of manic-depressive psychosis, the mechanisms involved were not known until the recent findings concerning linkage with markers on the X chromosome—color blindness and Xg blood group—in the families of bipolar depressives. The most recent findings in the genetic area were reported at the recent symposium by the American Psychopathological Association (56).

The internal environment model, which has been in the doldrums for awhile, is again in the forefront of scientific promise (113). The early work on the internal environment model proved disappointing because it dealt with peripheral blood flow and metabolic correlates in the brain. Furthermore, it was focused on discovering some unique toxic substance not present in normals. It was not until central synaptic mechanisms became the focus of attention that progress became apparent. Furthermore, the search for uniquely novel substances gave way to searching for inappropriate interconnections or interactions between chemically normal

---

components of the brain. Now at least three hypotheses seem promising. The first is the transmethylation hypothesis, which stipulates that the accumulation of methylated hallucinogenic substances underlies schizophrenic behavior. The second hypothesis postulates that the disturbance in central catecholamine synapses underlies the vulnerability in schizophrenia and is based on the similarity to schizophrenia noted in the toxic psychosis induced by amphetamine. Apparently release of dopamine in the brain is the culprit. The third hypothesis postulates that monoamine oxidase which is markedly reduced in the platelets of schizophrenics may serve as a marker for schizophrenia (113).

Can these various scientific models be integrated in some way? If we could squeeze out the essence of the six models into a goblet, we would find that the common element in all of these models is the production of some kind of vulnerability in the mental patients. When they are subjected to sufficient stressful situations in life, this vulnerability produces an episode of illness. We know that depressive or manic episodes are time limited; that schizophrenic episodes are also time limited remains to be demonstrated, since some schizophrenics appear to remain ill indefinitely. From our prognostic literature we know that good premorbid s improve while poor premorbid s do not. If we make the assumption that both good and poor premorbid s have time-limited episodes from which they recover, it is possible to explain the paradox. Both good and poor premorbid s return to their premorbid level when the episode is finished. The good premorbid s recover and return to their premorbid status in life, but the poor premorbid s, even when the episode is over, still cannot cope. We are not even sure that the episode is finished; furthermore, because of their high vulnerability, they are catapulted into another episode soon after the first is ended. The problem facing clinical research is how to improve the poor premorbid s so that they will be able to cope with life's exigencies when an episode ends. Perhaps an investigation of good premorbid s who fail and of poor premorbid s who succeed may hold the key to the solution (293).8

Literature Cited


8 For other recent literature not reviewed by this author but highly recommended by colleagues, see (26, 52, 118, 232).

15. Groen, W. E., Naka Reduced range of se chronic nonparanoi J. Abnorm. Psychol.

662 ZUBIN et al

differences in psychiatric patients and normal controls. Proc. 77th Ann. Conv. APA 4:3-4


68. Garvey, D. S. 1975. Some measurement theory. J. Mich. Psychol. 28:


70. Friedman, H. I. 1975. Some measurement theory. J. Mich. Psychol. 28:


72. Garmey, N. St...
113. Kety, S. S. 1960. The genetics of schizotypy. 165-84
of schizophrenia. II. Science 129: 1590-96
113. Kety, S. S. 1975. Progress toward an understanding of the biological substrates of schizophrenia. See Ref. 56
120. Klerman, G. L. 1972. Clinical research in depression. See Ref. 294a, 165-93
134. Lewis, A. 1967. Problems presented by the ambiguous word "anxiety" as used in psychopathology. Israel Ann. Psychiat. 5:190-21
139. Lipton, M. 1972. Age differentiation in
BIOMETRIC APPROACH TO PSYCHOPATHOLOGY


175. Ibid 1972. Depressive typologies and response to amitriptyline. 120:147–156


197. Rappaport, M., Silverman, J., Hopkins, H. K., Hall, K. 1971. Phenothia-


zine effects on auditory signal detection in paranoid and nonparanoid schizophrenics. Science 174:723–25
198. Raskin, A. 1971. Drugs and depression subtypes. See Ref. 55a, 87–95
204. Richman, A. 1970. The use of case-records of psychiatric care in epidemiological research of mental disorders. See Ref. 90, 257–72
212. Salzinger, K. 1969. The place of operant conditioning of verbal behavior in psychotherapy. See Ref. 69, 375–95
225. Schneider, E. W., Carpenter, J. A.
1973. The idiopathic sign. Alcohol 23
228. Shackelford Psychiatry in the 1970s: A study of patients with... study, Q1 261–73
237. Schneider, E. W., Carpenter, J. A.
1973. The idiopathic sign. Alcohol 23


250. Spitzer, R. L., Endicott, J. 1975. Psychiatric rating scales in the evaluation of psychiatric treatment. See Ref. 69a


257. Steinman, A. R. 1944. Reaction time to change compared with other psychophysical methods. Arch. Psychol. 292


264. Sutton, S. 1973. Fact and artifact in the psychology of schizophrenia. See Ref. 88a, 197-213


270. Tasto, D. L., Suinn, R. M. 1972. Fear survey schedule changes on total and factor scores due to nontreatment effects. Behav. Ther. 3:275-78


275a. Waldbum, J. L. (Krieger), Sutton, S., Kerr, J. 1975. Shift of sensory modality and reaction time in schizophrenia. See Ref. 115, 167-76


282. Winokur, G. 1971. manic-depressive illness. 47-54


288. Wolpe, J., Lang, P. Psychometric scale for the assessment of the personality disorder. Behav. Res. Ther. 16

BIOMETRIC APPROACH TO PSYCHOPATHOLOGY 671


282. Winokur, G. 1971. The genetics of manic-depressive illness. See Ref. 55a, 47-54


