Photo

747

Library of Congress Cataloging-in-Publication Data

Advanced abnormal psychology / edited by Vincent B. Van Hasselt and Michel Hersen.
p. cm.
Includes bibliographical references and index.
ISBN 0-306-44547-6
1. Psychology, Pathological. 2. Mental illness--Treatment.
I. Van Hasselt, Vincent B. II. Hersen, Michel.
RC454.A325 1994
616.89--dc20

93-36092
CIP

ISBN 0-306-44547-6

2

Assessment Strategies VAN HASSELT

ARTHUR N. WIENS AND JAMES E. BRYAN

INTRODUCTION

It is likely that all of the authors who prepared chapters for this volume in advanced abnormal psychology, and all of its readers, will, upon reflection, realize that they have assumed some definition of normality and abnormality in human behavior. For example, some may have assumed that normality equates with "health" and that behavior is assumed to be within normal limits when no manifest psychopathology is evident. Others may have in mind an "ideal" of optimal functioning. Still others, including many psychologists, may think of normality in terms of "average" levels of functioning and consider both very low and very high scores on various assessment procedures as deviant. This approach to describing abnormality is based on the mathematical principle of the bell-shaped curve and describes variability of behavior within the context of the total group, and not within the context of one individual. We will leave it to other chapter authors to elucidate this definitional issue of normality and abnormality. We did want to call the reader's attention to the fact that there may be few absolute definitions of abnormality and few clear-cut boundaries between normal and abnormal.

We also want to remind the reader that "abnormal" behavior, or diagnosable mental disorder, is widespread in our society. The field of psychiatric epidemiology is the study of the pattern of occurrence of mental disorders and deals with the distribution, incidence, prevalence, and duration of psychiatric illness with respect to the physical, biological, and social environment in which people live. A recent definitive study of psychiatric epidemiology has been conducted and is being analyzed by Darrel Regier and his associates at the Division of Biometry and

ARTHUR N. WIENS AND JAMES E. BRYAN • Department of Medical Psychology, Oregon Health Sciences University, Portland, Oregon 97201-3098.

Advanced Abnormal Psychology, edited by Vincent B. Van Hasselt and Michel Hersen. Plenum Press, New York, 1994.

Epidemiology of the National Institute of Mental Health. One objective of their study is to provide the most accurate estimates of the incidence of alcohol, drug abuse, and other mental disorders in the United States.

The five sites for their studies were: New Haven, Connecticut; St. Louis, Missouri; Baltimore, Maryland; Durham, North Carolina; and Los Angeles, California. In each site, adults aged 18 years and over were selected from rural, suburban, and urban neighborhoods; the total sample size was 18,571. The NIMH Diagnostic Interview Schedule (DIS), discussed later in this chapter, was used as the case-identification instrument. A 1-month time frame of prevalence rates allowed an assessment of current illness and minimized recall problems. The authors concluded that 15.4% of the population 18 years of age and over fulfilled criteria for at least one alcohol, drug abuse, or other mental disorder during the period 1 month before interview. The prevalence rates of DIS disorders varied from 12.9% in St. Louis to 19.8% in Baltimore. Higher prevalence rates of most mental disorders were found among younger people (<45 years), with the exception of severe cognitive impairments. Men had higher rates of substance abuse and antisocial personality, whereas women had higher rates of affective, anxiety, and somatization disorders.

Rates for any DIS disorder covered an increase from 15.4% for a 1-month prevalence, to 19.1% for a 6-month period, and to 32.2% for a lifetime prevalence (Regier, Boyd, Burke, Rae, Myers, Kramer, Robins, George, Karno, & Locke, 1988).

The point that we want to make in these introductory comments is that the purview of abnormal psychology is very broad and that it encompasses many different people at various stages in their lives.

Referrals for Psychological Assessment

To give the reader of this chapter a further look at how assessment in abnormal psychology is practiced, we will review aspects of our own clinical practice in a medical psychology clinic that is located in a health sciences university that includes many different outpatient clinics, two hospitals, and several psychiatric inpatient wards. The clinic is also the setting for a residency training program in medical psychology.

Faculty clinicians and residents in medical psychology respond to referral requests for assessment of patients on the psychiatric inpatient wards, on various inpatient medical wards, and from many of the fifty, or so, outpatient clinics; and to requests for assessment that are self-initiated by persons from the community.

Consultation requests for inpatient psychiatry services provide a good example of the multiplicity of assessment approaches that may be used in acute-care situations. Within our own clinic we have found that such consultation requests most often involve difficult diagnostic questions, where clarification of diagnosis and related cognitive and/or emotional symptoms can have significant bearing upon treatment and discharge placement decisions.

While this has been a long-established role of psychologists in hospitals, relatively little has been written about the effectiveness of assessment data in improving diagnosis and treatment. Recently, Zacker (1989) used a quasi-experimental approach to examine the impact of psychological assessment on diagnostic outcome in a series of 70 hospital referrals in a community mental health center. He found high concordance between the diagnoses of those based on psychological

assessment and those made by the referring clinicians. Concordance improved after the assessment findings were reported; while the rate of agreement was 71% at the time of admission between the psychologist and attending clinician, it improved to 94% at discharge (using four broad diagnostic categories). Assessment findings contributed to change of diagnosis in a significant number of these cases. As the psychologists' conclusions were accepted in almost every instance, specialized assessment information was clearly highly valued.

Such referrals to our clinic typically involve cases where information about cognitive performance, intellectual level, and personality style help the hospital team conceptualize diagnosis and treatment approach. In a series of 30 recent consultation requests over a 4-month period, we found that 83% asked for assistance with diagnosis. The others involved cases where diagnosis was already clearly established (e.g., mental retardation and at least moderate-stage dementia), and assistance was requested with behavior management and help in identifying sources of recent increases in agitation and disruptive behavior.

Cognitive and neuropsychological assessment was employed in 63% of these referrals, involving differential diagnostic questions such as schizophrenia vs. drug-induced encephalopathy; presence of mental retardation and/or schizophrenia; early-onset dementia vs. schizophrenia; dementia secondary to HIV and/or substance abuse; and degree of depression.

Personality measures, both objective (e.g., MMPI, SCL-90-R) and projective (e.g., Rorschach, Thematic Apperception Test, Incomplete Sentences), were used in 73% of the referrals, which included questions of: major depression and/or presence of personality disorder; schizophrenia vs. major depression with psychotic features; personality disorder and/or posttraumatic stress disorder.

Many cases also involved both types of measures, for several purposes. Personality measures were routinely employed in inpatient neuropsychological evaluations, for example, to determine both the types of symptomatology that the patients were reporting (e.g., were these consistent with neurological and/or psychotic dysfunction?), and the extent of their emotional distress. Cognitive and intellectual measures were similarly employed in schizophrenia-related assessment. In these cases, WAIS-R performance was qualitatively analyzed (e.g., assessing looseness of association or unusual linguistic errors on the open-ended verbal subtests). Visual–perceptual and organizational functioning was also assessed as required in the copy reproduction and recall trials of the Rey-Osterreith Complex Figure Test. Diagnostic conclusions in most cases thus involved a combination of quantitative and qualitative assessment.

Communication with the referring clinician before, during, and following the evaluation is emphasized routinely, and has further enhanced the value of formal test findings. At the outset, review of medical records and briefing with the clinician helps determine the selection of tests and analysis of results in regard to pertinent questions. Follow-up reporting to both the clinician and the patient helps coordinate test results with treatment and discharge planning and enhances collegial and collaborative relationships (Zacker, 1989).

Many different assessment/referral questions are presented by patients from medical wards or from other clinics in our health care center. Familiar diagnostic issues about depression, bipolar affective disorders, anxiety, panic attacks, and so on, are often presented. We also see many patients who are concurrently evaluated by psychologists and physicians because of the awareness that both

psychological and physical factors are intertwined in the symptoms and distress the patient experiences, e.g., patients presenting with fibromyalgia, chronic fatigue syndrome, chronic pain. Patients are also seen for evaluation of stress reactions and stress management or temper outbursts and anger management. Many patients experience a decrement in cognitive abilities: Complaints about memory impairment are often received and are assessed. The cognitive sequelae of head injury are evaluated and the progressive dementia that may accompany HIV/AIDS is often monitored by referral for serial psychological evaluation.

Other groups of patients that we see in our clinic are those who have been exposed to industrial or other toxins and fear that they may have suffered central nervous system damage. Still other patients feel that their psychological or physical impairments are so severe that they are disabled from any gainful employment; both patients and governmental agencies may request psychological assessment to help evaluate such claims.

There are also referral requests that arise out of new and innovative health care procedures, and the psychologist may be called upon to devise assessment procedures and protocols to address questions that may be entirely new. The scientist-professional education and training background of the psychologist is often invaluable to devise innovative assessment techniques and establishing the protocols that will allow systematic evaluation of the reliability and validity of the new techniques, and even the validity of the new health care interventions that are being used. Some of the new interventions include organ transplantations that involve the psychological assessment of a patient's capacity to withstand the rigors of the procedures, as well as the patient's ability to comply with the medical management regimens that follow. Another important question has to do with the quality of a patient's life after a particular medical intervention; this will be described further in a later section of this chapter. As the reader of this chapter has no doubt already surmised, we view assessment in abnormal psychology to be very wide-ranging. The successful practitioner in this area needs to be broadly and intensively educated and trained.

Practice Data Management

An important assessment strategy in abnormal psychology or, in any other aspect of psychological practice, is to establish the reliability and validity of the psychologist-assessor. To this end the successful psychologist will establish those practice data management procedures that will allow a detailed description of patients seen, verification of diagnosis or other assessment conclusions, and ultimately the determination of those patient groups that a given psychologist can successfully diagnose and treat. Practice in psychology is becoming increasingly specialized, and specific expertise in an area of practice will have to be demonstrated in the future.

In our own clinic we have developed a clinic management program or clinic appointment activity report that allows us to see how many of the patient appointments that were scheduled were kept, not kept, or canceled. We can describe the age and sex distribution of our patients, the procedures that were completed with them, and the diagnoses assigned to them. We can examine the data for all of our clinicians combined or for a given clinician; for example, each of the residents

in our program can have a printout report of the patients that he or she has seen. This activity report has administrative, professional, and educational uses. It allows each resident to track and have a record of the assessment procedures done and the diagnoses of the patients seen. This also creates the possibility of assigning certain kinds of patients to remedy deficits in training. The reports create documentation that the resident can later use when asked to verify the nature and extent of training experiences.

To illustrate again the variety of assessment demands in abnormal psychology, we can note that over a 3-month period our residents saw about an equal number of males (49.2%) and females (50.8%). Approximately 10% of the patient appointments involved children and adolescents under 21 years of age, and 8.6% of the appointments involved patients 65 years of age or older. About 17% of the appointments involved psychiatric inpatients and about 16% involved inpatients from various medical wards. Many of those seen were outpatients referred either from other clinics or from the community. For administrative purposes, we also note that only about 10% of our patients were nonsponsored (i.e., the majority of our patients were covered by commercial insurance carriers or other health care contracts).

Approximately one-half of our clinic patient appointments were for treatment procedures, e.g., couples' therapy, individual therapy, tension-pain therapy, or brief visits. The assessment procedures that were completed included:

- Parenting Evaluation and Report
- Consultation Interviews
- Intellectual Evaluations
- Intellectual-Personality Evaluations
- Neuropsychological Evaluations (complete)
- Neuropsychological Evaluations (partial)
- Personality Evaluations
- Psychophysiologic Evaluations
- Behavioral Evaluation and Report
- Intellectual and Developmental Evaluation and Report

Intellectual and Behavioral Evaluation and Report. The most common procedure done in our clinic for the time period reviewed here was the Neuropsychological Evaluation (partial). In addition to a Clinical Interview and a Psychological/Social History Questionnaire, the modal psychological test battery in this evaluation includes the Wechsler Adult Intelligence Scale—Revised, the Wide Range Achievement Test—Revised-2, the California Verbal Learning Test, The Rey-Osterreith Complex Figure Test, Trail Making Test—Parts A & B from the Halstead-Reitan Neuropsychology Test Battery, the Hopkins Symptom Check List (SCL-90-R), and often the Minnesota Multiphasic Personality Inventory.

It is also useful to note the diagnoses assigned to the patients seen to illustrate the range of referral questions that can be inferred from these diagnoses. The assigned diagnoses in this period were as follows:

- presenile dementia, uncomplicated
- senile dementia with delirium
- alcoholic psychoses
- amnesic syndrome
- dementia in conditions classified elsewhere
- other specified organic brain syndromes (chronic)

- paranoid type schizophrenia
- schizo-affective type schizophrenia
- unspecified schizophrenia
- major depressive disorder, single episode
- major depressive disorder, recurrent episode
- bipolar affective disorder, unspecified
- paranoid states
- anxiety state, unspecified
- panic disorder
- neurotic depression
- somatization disorder
- affective personality disorder, unspecified
- schizoid personality disorder, unspecified
- schizotypal personality
- histrionic personality disorder
- dependent personality disorder
- antisocial personality disorder
- borderline personality
- other personality disorders
- alcohol dependence syndrome
- other and unspecified alcohol dependence
- barbiturate and similar sedative/hypnotic dependence
- cannabis abuse
- other, mixed, or unspecified drug abuse
- tics
- tension headache
- brief depressive reaction
- prolonged depressive reaction
- specific academic or work inhibition
- adjustment reaction with anxious mood
- adjustment reaction with mixed emotional features
- other adjustment reaction with emotion disturbance
- adjustment reaction, emotion and conduct disturbance
- adjustment reaction with physical symptoms
- unspecified adjustment reaction
- frontal lobe syndrome
- organic personality syndrome
- postconcussion syndrome
- other brain damage, nonpsychotic mental disorders
- unspecified brain damage, nonpsychotic mental disorders
- socialized conduct disorder
- overanxious disorder
- unspecified delay in development
- psychic factors associated with diseases classified elsewhere
- mental retardation: mild, severe, profound
- other specified mental retardation
- migraine headache

From the list of diagnoses found for these patients it seems clear that many different assessment strategies would have to be used in assessing them.

Assessment Strategies

Interviewing

The Clinical Interview

The topics to be covered in an initial clinical interview are relatively consistent from one clinician to the next. The general objective is to carefully obtain a history that can be the foundation for the diagnosis and treatment of the patient's disorder. More specific objectives of the clinical interview are to understand the individual patient's personality characteristics, including both strengths and weaknesses; to obtain insight into the nature of their relationships with those closest to them, both

past and present; and to obtain a reasonably comprehensive picture of the patient's development from the formative years to the present.

In preparing a written record of a clinical interview, most clinicians begin by presenting *identifying information*, such as the patient's name, age, marital status, sex, occupation, race, place of residence and circumstances of living, history of prior clinical contacts, and referral and information sources. The *chief complaint*, or the problem for which the patient seeks professional help, is usually reviewed next and is stated in the patient's own words or in the words of the person supplying this information. The intensity and duration of the presenting problem is noted, specifically the length of time each symptom has existed and whether there have been changes in quality and quantity from a previous state. It is also useful to include a description of the patient's appearance and behavior. In reviewing a *present illness* or presenting problem, the clinician looks for the earliest and most disabling behavior or symptoms and for any precipitating factors leading to the chief complaint. Often the precipitating or stress factors associated with onset of symptoms may be subtle and require the clinician to draw on knowledge of behavior and psychopathology to help with inquiry regarding relevant life change events. The clinician should also report on how the patient's problems have affected his or her life activities. It is important to review *past health history* for both physical and psychological problems—for example, physical illnesses that might be affecting the patient's emotional state. Prior episodes of emotional and mental disturbances should be described. The clinician also needs to inquire about and report prescribed and nonprescribed medication and alcohol and drug use. Possible organic mental syndromes must be noted. *Personal and social history* usually includes information about the patient's parents and other family members and any history of psychological or physical problems. The account of the patient's own childhood and developmental experiences may be detailed. Educational and occupational history are noted as well as social, marital, military, legal, and other experiences. The personal history should provide a comprehensive portrait of the patient independent of his or her illness (Siassi, 1984). The mental status examination is also included in the initial clinical interview but will be reviewed separately in the next section. The section of the interview on initial impressions or *findings* should include deductions made by the clinician from all sources available to this point about the patient's past history, description of the present problems, and results of the clinician's examination as determined from the mental status examination, results of psychological testing, contributions of family members and significant others, and so on. Finally, *recommendations* are presented about what kind of treatment the patient should receive for what problems and target symptoms.

Mental Status Examination

The mental status examination is reviewed under the following headings: general appearance and behavior; mood, feelings, and affect; perception; speech and thought; sensorium and cognition; judgment; insight; and reliability.

An example of a mental status examination report that was printed from responses recorded in a structured interview (Harrell, 1984) is presented below:

> Ms. Doe was generally cooperative with the interviewer although her specific interactions were defensive. Level of consciousness during the interview was

unimpaired. She was not under the influence of alcohol or drugs. Ms. Doe was oriented to time, place, person, and situation. No apparent deficits were evidenced in attention and concentration. Comprehension of simple commands was unimpaired. There *was* evidence of impairment in short-term memory. No indications of amnesia were present. Current intellectual level appeared to be average and fund of information was below average. Current intellectual functioning appears to be consistent with that evidenced prior to onset of the present condition. Abstract thinking appeared intact. No impairment was evidenced in simple computational skills. There was no evidence of specific neurological impairment. Examination of perceptual processes did not reveal any illusions, hallucinations, or other perceptual dysfunctions. No unusual aspects of thought content were noted. There was no evidence of phobias. The predominant mood during the evaluation was moderate anger, which was consistent with thought content. Secondary moods included mild depression. Generally, affective reactions were appropriate to status or complaints. Appropriate variability in affective reactions was observed. There was no evidence of significant cyclic mood changes. There was no indication that Ms. Doe is currently at risk for suicide. Current risk of danger to others appears to be low. No current self-destructive behavior patterns were identified. Level of impulse control was estimated to be limited and judgment generally appeared to be below average. Level of insight was characterized by some awareness of problems with some denial.

Communication between Doctor and Patient

Although a treatise on the importance of the doctor–patient relationship is beyond the scope of this chapter, it is necessary to point out that a clinical interview, or mental status examination, cannot be conducted with validity unless reasonable rapport is established and the doctor and patient are listening to each other.

In a study of more than 1000 encounters between internists and patients, Beckman and Frankel (1984) have reported that most people are interrupted by their physicians within the first 18 seconds of beginning to explain what is wrong with them. This practice often prevents people from completing the purpose of their visit.

Typically, people go to their physicians with about three concerns, and the most troubling complaint is not always presented first. No relationship was found between the order of presentation and the importance of the complaint. This finding challenges the prevailing hypothesis that the first complaint is the most important. Once interrupted early in the encounter, patients rarely return to any additional concerns. The researchers found no differences between male and female doctors in the tendency to interrupt and control the interview. They also found that an encounter averages about 15 minutes in the United States and about half that in Great Britain.

The doctor–patient encounter can be made more useful if patients think beforehand about what they want to say and get out of the visit and take more control of the interview. It appears that older people are less willing to assert themselves in this manner than younger people, and thus they may be at particular risk for not having their concerns heard. Basically, patients want clinicians who will

work with them and who will understand them, and a very important strategy in assessment in abnormal psychology is the establishment of rapport or a working relationship between doctor and patient.

Structured Interviews

A major source of unreliability in diagnosis in abnormal psychology is the variability of information about a patient that is available to a given clinician. For example, some clinicians may talk with patients' families and others may not. Similarly, some clinicians may ask questions concerning areas of functioning and symptoms and other clinicians may not.

To deal with such information variance, psychologists and psychiatrists have developed structured clinical interviews that reduce that portion of the unreliability variance based on different interviewing styles and coverage. The structured clinical interview is used routinely in clinical research and increasingly in daily clinical patient examinations. A structured clinical interview essentially outlines a list of target behaviors, symptoms, and events to be covered, and some guidelines or rules for conducting the interview and recording the data. Interview schedules vary in that some offer only general and flexible guidelines and others have strict and detailed rules (i.e., some are semistructured and others are highly structured). With the latter, wording and sequence of questions, recording responses, and rating responses are all specified and defined. The interviewer may be regarded as an interchangeable piece of the assessment machinery. Clinical judgment in eliciting and recording information is minimized and, given the same patient, different interviewers should obtain the same information. The impact of computers in standardized interviewing also appears decisive, in that they allow for efficient retrieval of information. Computers can also be used to apply an algorithm to yield reliable diagnoses from raw data.

Diagnostic Interview Schedule (DIS)

The DIS (Robins, Helzer, Croughan, & Ratcliff, 1981) is a fully structured interview schedule designed to enable clinicians to make consistent and accurate DSM-III psychiatric diagnoses. It was designed to be administered by persons not professionally trained in clinical psychiatry or psychology, and all of the questions and the probes to be used are fully explained. It reminds interviewers not to omit critical questions and presents well-tested phrasing for symptoms that are difficult to explain or potentially embarrassing to patients. Questions about symptoms cover both their presence or absence and severity (e.g., taking medication for the symptoms, seeing a professional about the symptom, and having the symptom significantly interfere with one's life). In addition, the interview ascertains whether the symptom was explained entirely by physical illness or injury or as a complication of the use of medication, illicit drugs, or alcohol. The age at which a given diagnostic symptom first appeared is also determined, along with the most recent experience of the symptom. These questions are designed to help determine whether a disorder is current (i.e., the last 2 weeks, the last month, the last 6 months, or the last year). Demographic information, including age, sex, occupation, race, education, marital status, and history of treatment, is also determined.

Current functioning is evaluated by ability within the last 12 months to work or attend school, maintain an active social life, act as head or cohead of a household, and get along without professional care for physical or emotional problems.

Aside from a few open-ended questions at the start of the interview to allow the interviewee the opportunity to voice the chief complaint and to give the interviewer some background for understanding answers to close-ended questions, the interview is completely precoded. Symptoms assessed by the computer are precoded at five levels: (a) negative, the problem has never occurred; (b) present but so minimal as to be of no diagnostic significance; (c) present and meets criteria for severity, but not relevant to the psychiatric diagnosis in question because every occurrence resulted from the direct or side effects of prescribed, over-the-counter, or illicit drugs or alcohol; (d) present and meets criteria for severity but not relevant to the psychiatric diagnosis in question because every occurrence resulted from medical illness or injury; and (e) present, meets criteria for severity, and is relevant to the psychiatric diagnosis under consideration.

The DIS has been translated into different languages, and its use is now underway, or planned, in about 20 different countries. Cross-national comparisons in psychiatric and psychological epidemiology are possible due to the growing number of population surveys in various countries that have used the DIS. Similarly, cross-cultural surveys of anxiety disorders and prevalence, and symptomatic expression and risk factors in alcoholism have been planned.

Computerization of the DIS makes direct patient administration possible either in its entirety (18 sections) or one section at a time. The computer printout lists all DSM-III diagnoses for which the patient meets criteria. It also presents additional information about each diagnosis including the recency of symptoms, duration, and age of onset. In addition, the printout lists for the clinician what other diagnoses must be ruled out before this diagnosis can be assigned according to the DSM-III hierarchy. The diagnostic categories that are surveyed include the following:

- Tobacco Use Disorder
- Somatization Disorder
- Panic Disorder
- Generalized Anxiety Disorder
- Phobic Disorder
- Depression
- Manic Episode
- Schizophrenia
- Anorexia Nervosa
- Bulimia
- Alcohol Abuse and Dependence
- Obsessive Compulsive Disorder
- Drug Abuse
- Conduct Disorder
- Psychosexual Dysfunction
- Antisocial Personality Disorder
- Pathological Gambling

We have been using the Computerized-DIS (Blouin, 1985) in our clinic for some time. As part of our assessment strategy we wanted to include a procedure that clinician/researchers in other settings could follow if they wished to replicate our clinical data. We have been using the DIS routinely in examination of fibromyalgia patients. This is an example of a particular patient population in which we assumed that psychological factors had an important role either as an etiological factor in the development of the illness or as a consequence of suffering from it.

To date, after more than 100 referrals, no patient has yet been unable to complete the DIS. Our patients appear to be willing, and perhaps to welcome the opportunity, to respond to the various questions about their health status and

history. Time at the computer screen to complete the DIS has varied from about 45 minutes to more than 2 hours, depending on how many different question branches the patient's responses would yield. With knowledge of the diagnostic presence of one or another DSM-III disorder, it is possible to tailor the patient's treatment program to be more comprehensive by taking psychological factors/needs into account. First developed to help in making DSM-III diagnoses, the DIS was later modified to assist in making DSM-III-R diagnoses. It will undoubtedly be modified again to assist in making DSM-IV diagnoses (American Psychiatric Association, 1993).

Psychological/Social History Report

Another standardized data collection questionnaire that we use routinely is the Psychological/Social History Report (Rainwater & Coe, 1988). The questions are in a multiple-choice response format that permits computer scoring of the responses with a narrative printout of the results. Question categories include family/developmental experiences, educational experiences, employment experiences, military history, alcohol and drug use history, medical history, marriage history, diet, psychological history, and presenting problems. As with many questionnaires, the patient is asked to respond to many more questions than a face-to-face interviewer might have patience to pursue. Responding to all of the questions prevents overlooking a critical problem area in the patient's life, that might later turn out to be an important assessment omission.

Research has indicated (Young, O'Brien, Gutterman, & Cohen, 1987) that structured interviews increase, by a factor of two to one, the number of clinical observations (e.g., number of problem areas) and the amount of relevant patient information that is recorded. Clinicians using structured interviews tend not to be limited to the presenting symptoms in their diagnostic formulations; their results have higher reliability. Interviewers using structured interviews consider themselves equally as emphathic as when using free-flowing interviews. With practice, they can use structured interviews with increasing efficiency, so that this method requires about the same amount of time as traditional clinical interviews.

AutoSCID II

The AutoSCID II (First, Gibbon, Williams & Spitzer, 1991) is a computer-administered version of the "Structured Clinical Interview for DSM-III-R Personality Questionnaire" (SCID II PQ). It has been designed to assist in the assessment of personality disorders and can be used to collect diagnostically relevant historical data directly from the patient using SCID II PQ questions. The clinician can be prompted by the responses the patient has made to the screening questions to inquire further about evidence for the different personality disorders. It can be used to screen for the presence of adult Axis II disorders, as identified in the DSM-III-R.

This diagnostic approach to personality disorder views personality traits as enduring patterns of perceiving, relating to, and thinking about the environment and oneself, which are exhibited in a wide range of important social and personal contexts. It is only when personality traits are inflexible and maladaptive and cause either significant functional impairment or subjective distress that they constitute personality disorders (DSM-III-R: American Psychiatric Association, 1987). To be

rated as present the described characteristic must show evidence of being pathological, persistent, and pervasive. Pathological characteristics must be beyond those experiences that one would expect to see in nearly everyone; for example, social anxiety would have to be clearly extreme. To be diagnosed, a characteristic should have been present over a period of at least 5 years. The characteristic should also be apparent in a variety of contexts, such as at work and at home, or in different relationships.

Computer-Administered Interviews

Computers have long played a significant role in assessment. Much modern test construction has been dependent on the availability of computing resources. As test administration itself became more feasible with the advent of microcomputers, one of the questions raised concerned the comparability of data obtained with traditional paper-and-pencil administration and computerized administration. Lukin, Dowd, Plake, and Kraft (1985) obtained no significant differences between scores on measures of anxiety, depression, and psychological reactance across administration formats. Most important, while producing results comparable to the pencil-and-paper assessment, the computerized administration was preferred to the pencil-and-paper administration by 85% of the subjects.

More recently, Choca and Morris (1992) compared a computerized version of the Halstead Category Test to the standard projector version of the test using neurologically impaired adult patients. Every patient was tested with both versions and the order of administration was alternated. Results indicated that difference in mean number of errors made between the two versions of the test was not significant. The scores obtained with the two versions were seen as similar to what would be expected from a test-retest administration of the same instrument. The authors note that one advantage of the computerized version is that it assures an error-free administration of the test. Secondly, the computer version allows the collection of additional data as the test is administered, such as the reaction time and the number of perseverations when a previous rule is inappropriately used. Finally, it may be eventually possible to show that promptings from the examiner do not make a significant difference in terms of the eventual outcome. If this were the case, the computer version would have the added advantage of requiring a considerably smaller time commitment by the examiner (Choca & Morris, 1992, p. 11–12).

There is evidence (Giannetti, 1987) that automated self-reports have advantages for both clinical practice and research. Patients accept and enjoy responding to online computerized questionnaires and frequently prefer them to clinical interviews or paper-and-pencil questionnaires. Even chronic and disturbed inpatients can answer computer-presented questions without assistance. There are indications that respondents are more likely to report socially undesirable behavior to a computer (e.g., reporting greater alcohol consumption to computers than to interviewers). Self-report and interviewer-collected history data show high agreement. Finally, it may be cost saving to complete interviews by computer rather than by traditional means.

Adams and Heaton (1987) called attention to a further administrative/research role of computers in clinical practice: creating and maintaining an informational database. This database might include information concerning patient demo-

graphics, referral sources, historical data, criterion test results (e.g., brain tests), psychological test findings, and clinical outcome. Such information is valuable in documenting the sources of patients, their demographic and base-rate profiles, the relationship of neuropsychological tests to other results, and the impact of testing, or other services, on patient outcome. Such data are of importance in quality assurance and in evaluation research. External reviewers and third-party agencies increasingly request data showing accuracy of diagnosis and relationship to hospital/clinic utilization, more appropriate care, and improved outcome. Practice data are important to have, given the current climate in health services delivery. Once this view is accepted, it follows that the optimal way to gain control of the quality and accuracy of such data is to implement one's own system to generate them.

Information from Family Members and Other Collaterals

The clinical usefulness of psychological test results is determined by its relation to the person's functioning in everyday life. Literature as to the "ecological validity" of psychological assessment consistently compares psychometric test data with other measures of real-world functioning (e.g., Baird, Brown, Adams, & Schatz, 1987). Information obtained from others who live with and know the person is essential to understanding the nature of the condition and predicting its long-term effects. Such information from family members and others closely involved with the person ("collaterals") can serve many purposes.

In diagnosis, other persons can help describe changes in the person's functioning, in terms of deterioration of general social roles (e.g., employment status and quality, domestic responsibilities, recreation and activity level), or global change in personality and emotional style. They can also help specify clinically significant symptoms that may not emerge through psychometric methods or that the individual may not recognize or acknowledge (e.g., the "positive spouse sign" is familiar to interviewers of early-stage dementia patients, when the spouse points out significant deficits that the patient minimizes or fails to mention).

In terms of ongoing behavior management, family members may be able to alert the clinician to specific problematic or at-risk behaviors requiring rapid attention. The degree of discrepancy between family members' and patients' ratings of functioning on the same measure can be informative. Discrepancies may point to the patients' lack of awareness of symptoms, as well as to possible hypervigilance and overinvolvement of the relatives. As a source of outcome data, the same measures can be administered to relatives over time, before and after treatment and at successive intervals afterward. Family members may be able to describe the effectiveness of interventions and provide suggestions to better apply methods to their own situation.

Family rating methods typically address global estimation of adjustment and role function as well as more specific important behaviors and cognitive/emotional symptoms. They represent a systematic means of gaining highly relevant dispositional information about the person, to complement clinical psychometric information or supplement it when complete testing may not be possible. Issues of psychometric reliability and validity are central to their meaningfulness and usefulness. These are commonly in the form of self-administered paper-and-pencil measures as well as structured face-to-face or telephone interviews. Some of the more commonly employed measures will be described here.

Self-Administered Measures

These are completed by the collateral member alone, and typically involve scoring level of functioning on a variety of general and specific items. The *Katz Adjustment Scale-Relatives Form* (Hogarty & Katz, 1971) is a measure of quality of life for both patient and relative. It has been used in a wide variety of studies involving health care conditions that compromise everyday functioning. Relatives rate two areas: (1) the person's current level of performance across a wide variety of daily social role functions; and (2) their estimation of how well they feel the person is meeting what they *expect* of them in that capacity. This provides both a measure of level of function as well as of discrepancy from the relative's expectations.

The *Cognitive Behavior Rating Scales* (Williams, 1987) address common sequelae of neurological conditions. The raters indicate their observations of the presence and degree of severity of a range of emotional and behavioral symptoms, neurological signs, and cognitive deficits. These are organized under nine scales entitled Language Deficit, Apraxia, Disorientation, Agitation, Need for Routine, Depression, Higher Cognitive Deficits, Memory Disorder, and Diffuse Dementia.

Briefer and more behaviorally specific rating forms have been developed with particular purposes in mind. An example is the *Patient Competency Rating Form* developed by Prigatano (1986), as part of a comprehensive neuropsychological rehabilitation program. This measure rates observations of the patient's ability to perform a variety of daily living skills, from hygiene to cooking, with the degree of difficulty presented by emotional and cognitive symptoms. The *Disability Rating Scale* (Rappaport, Hall Hopkins, Belleza, & Cope, 1982) is a similar brief measure of self-care, cognitive level, and occupational functioning designed for head-injured patients in clinic and at home.

Family Interview Measures

Other family rating systems are placed in the context of a structured interview. This permits rating of the same sorts of areas noted above, but in a more qualitative and content-oriented format of contact with the family member. The *Vineland Adaptive Rating Scales* (Sparrow, Balla, & Cicchetti, 1984) are among the most widely used in this regard. They allow highly useful information to be obtained about the adaptive functioning of impaired and mentally retarded persons, which cannot be as clearly described by test data, particularly for severely impaired persons. They are scaled by norm group and age equivalent in four broad domains: Communication, Socialization, Motor Skills, and Maladaptive Behavior. This format allows an estimation of general level of function in each area, as well as the drawing of a profile of adaptive strengths and weaknesses in the person's daily living.

Other behaviorally oriented interviews include the *Social Behavior Assessment Schedule* (Platt, Weyman, Hirsch, & Hewett, 1980), designed to assess the extent of dysfunction of schizophrenic patients, and its impact upon persons who live with them. In this format, the family member is asked to rate degree of severity in a variety of behavioral and psychiatric-symptom-related areas, along with the extent to which they find those symptoms distressing to them. This has been applied to other populations, such as the head injured, and used as a measure of subjective

burden involved in caregiving for such persons at home (Bryan & Strachan, 1992). *The Burden Interview* (Zarit, Orr, & Zarit, 1985) is a brief, symptom-focused rating system also designed to assess the degree of distress associated with caring for dementia patients.

Observed Interactions

Parent–Child Interaction

The opportunity to observe individuals interacting with each other often provides a great deal of information about each of the individuals as well as their relationship. Observed interaction is often of critical importance in evaluating children, whose behavior may in large part be a reflection of the stimulus values and reinforcement behaviors of the parents. Robinson and Eyberg (1981) asserted that direct observation is a critical component of clinical child assessment and described an observational system to do such assessment. They described a study in which they standardized and validated the Dyadic Parent–Child Interaction Coding System (DPICS), a comprehensive observational system for conduct problem children. Both parent and child behaviors are observed and coded. Each parent (i.e., mother and/or father, if available) was observed in two 5-minute interactions with each child in a playroom. There were two types of interaction. In the child-directed interaction, the parent was instructed to allow the child to choose any activity and to play along with him/her. In the parent-directed interaction, the parent was instructed to select an activity and keep the child playing according to the parent's rules.

Interrater reliability was assessed; the mean reliability coefficient for parent behaviors was 0.91 and for child behaviors, 0.92. Validity was investigated by examining differences between normal and conduct problem families. Parents of conduct problem children made more critical statements and direct commands and gave fewer descriptive questions than did parents of normal children. In addition, the conduct problem parents gave a higher percentage of direct commands to their children than did normals. The conduct problem children demonstrated more whining, yelling, and noncompliance than normal children. For example, the average normal child noncomplied 6.1 times, whereas the average conduct problem child noncomplied 14.2 times during 10 minutes of observation. The DPICS correctly classified 94% of families and predicted 61% of the variance in parent report of home behavior problems.

Robinson and Eyberg (1981) suggest that continuous recording contributes to validity and utility by providing a complete account of all behavior, and that it allows data to be collected in less time than typically is required by interval sampling methods. They also note that the structure of the situations permits both the parent and the child to proceed naturally under varying degrees of parental control, thus maximizing the possibility of observing interactional dysfunction in conduct problem families. The authors explicitly acknowledge that characteristics of the parent as well as those of the child contribute to the diagnosis of conduct problem. Finally, they point out that their observational procedure can be used serially to guide the course of treatment and to document treatment change.

The DPICS has been used clinically in different assessment situations. As

already noted, it can be used to assess conduct problems between parent and child and to monitor change with treatment intervention. It has also been used in clinical assessment when there are clinical/forensic questions, such as determination of child custody and termination of parental rights. The generation and availability of empirical observation data can be useful in such decision making.

Inpatient Ward Observation Methods

Among the most important sources of information about patients is observation of how they function while in the hospital. In order to sensibly interpret and organize the inpatient observation information, structured rating scales have been developed. These typically place high value upon being brief and convenient, while also seeking to adapt acceptable levels of reliability and validity. The best known of these methods have evolved from inpatient psychiatry. They tend to be structured around diagnostically significant behaviors and related observable symptoms. Clinically, they are used to help diagnose and track changes in patients' mental state and functional status over time. They are also used to help establish the validity of diagnostic systems through research on large groups of patients, making use of the wealth of available information from patients within a controlled setting.

The most common inpatient psychiatric rating methods have been in use for decades, including the *Brief Psychiatric Rating Scale* (BPRS) (Overall & Gorham, 1962), the *Present State Examination* (Wing, Cooper, & Sartorius, 1974), and the *Nurses' Observation Scale for Inpatient Evaluation* (NOSIE) (Honigfeld, Gillis, & Klett, 1966). They enjoy widespread use, with continuing refinement of psychometric and diagnostic discriminatory properties. These typically rate both global level of functioning and specific descriptive subscales.

Recent modifications of the BPRS include development of an 18-item format in which the clinicians rate on a scale of very mild to severe, items described by one or two sentences. The BPRS-18 includes the following six scales: Anxiety-Depression, Lack of Energy, Thought Disturbance, Hostility-Suspiciousness, and "Schizophrenia" (including unusual thought content, hallucinations, blunted affect, and emotional withdrawal). Hafkenscheid (1991) found that the thought disturbance and schizophrenia scales and the global scale showed adequate reliability and discriminatory power. The NOSIE has also been found to be convenient and sensitive to clinical change. Recent studies have reported good reliability and identified six main factors: Social Competence, Social Interest, Personal Neatness, Irritability, Psychoticism, Retardation, and Depression (Dingemans, Bleeker, & Frohn-DeWinter, 1984).

These scales have been adopted in current research on "positive" and "negative" symptoms in diagnostic subtypes of schizophrenia. Positive symptoms involve abnormal and maladaptive functions such as hallucinations, bizarre behavior, and disturbed thought. Negative symptoms relate to the absence or lack of aspects of normal functioning and include poor initiative, social withdrawal, flat affect, and impoverishment of thought and speech. Dingemans (1990) found that both the BPRS and NOSIE contributed to reliable identification of positive symptoms, although negative symptoms were less consistently measured. Greater use of these scales to track patient progress and operationalize diagnostic constructs is expected.

Quality of Life Assessment

Advances in the effectiveness and expense of clinical health procedures have raised pressing questions about how and when they should be applied. The benefits versus costs of medical procedures are the focus of crucial economic and biomedical ethical decision making. The concept of *quality of life* has evolved as the measure of worth to be balanced against the efforts, costs, and risks of intervention methods. The impact on the quality of life of recipients is addressed at many levels, from political/social resource distribution, to measures of the effectiveness of drug, surgical, or psychological treatment, to clinical contact with individual patients involving the ongoing assessment of their progress and response. Quality of life is further used as an outcome measure in individual and epidemiologic studies, and as the basis for establishing standard of care in many areas.

While a crucial construct, quality of life is also broad and vague. It is generally agreed that it can best be conceived as a multidimensional construct, an aggregate of distinguishable and closely related factors affecting response to the disease and/or treatment. Definitions of the construct typically include:

1. Physical aspects, including mobility, pain, and appearance
2. Psychological and emotional aspects, including cognitive/intellectual function, self-esteem, and subjective sense of well-being
3. Social aspects including role functioning, contact versus isolation, and reciprocation in relationships (Siegrist & Junge, 1989)

Shumaker, Anderson, and Czajkowski (1990) have further added productivity and intimate involvement with others as relevant dimensions. Spilker (1990) has proposed a system-based definition that couches the safety, efficiency, and cost of treatment in terms of outcomes of physical status, psychological well-being, social interactions, and economic status. The patient's values, beliefs, and judgments represent moderating factors also affecting outcome.

Given the broad definition and wide range of applications of this construct, its meaning is being established through specific uses and measurement strategies. Shumaker et al. (1990) recommend a hypothesis-testing approach, with emphasis upon the specific areas considered to be relevant to a particular treatment. "Quality of life" seems all-encompassing, while prediction about types of expected effects forces the examiner to consider which dimensions are most relevant to the clinical trial.

Considerable attention has been given to the psychometric foundations of measures of quality of life. Objective, reliable, and standardized measurement is especially important in such a value-laden area. As a result, most measures emphasize the types of specific dimensions mentioned above. Some approaches include all of them within a single instrument, while other authors advocate a battery of individual tests. To be useful and meaningful, such measures must meet standards of reliability and validity (in all forms). Further, they must be sensitive indicators of change, providing information about which dimensions are affected by treatment, in which directions, and at which times.

A variety of approaches have been employed in quality of life assessment. They include scaled observer ratings and triangulation of information from differ-

ent sources such as family members, co-workers, and physicians (Siegrist & Junge, 1989). The self-report questionnaire format is by far the most common, due to many advantages of data collection, cost-efficiency, and applicability to a range of patients. They are most effective when compared with other indices of physical, psychological, and social functioning, including ratings by physicians and others who know the person well.

Many self-report measures have been developed, the most common of which will be summarized as examples of the types of dimensions addressed and questions utilized. The *General Health Questionnaire* (Goldberg, 1979) is a 60-item scale developed as a screening measure of somatic symptoms, mood and affective states, subjective feelings of distress, and social interactions. The *Symptom Checklist-90-Revised* (Derogatis, 1983) is a similar 90-item rating scale of symptoms primarily in nine psychiatric categories, and yields both separate scale scores and a Global Symptom Index. The *Chronic Illness Problem Inventory* (Kames, Naliboff, Heinrich, & Schag, 1984), permits rating of perceived severity of limitation upon a wide range of areas of daily functioning among persons affected by debilitating illnesses. The *Quality of Life Scale* (Burkhardt, Woods, Schultz, & Tiebarth, 1989) is a brief, 16-item scale that yields a summary satisfaction score regarding broad areas of life. Specific scales have been developed for more circumscribed populations, such as the *Quality of Life Rating Scale* (Walker, Blankenship, Ditty, & Lynch, 1987), a nine-item Likert-scaled measure designed for clients within a head-injury rehabilitation program.

The *Quality of Well-Being Scale*, developed by Kaplan and Anderson (1990), is a functional health measure completed by the clinician. It consists of three scales that focus upon elementary aspects of daily functioning: social role activity, physical activity, and mobility. These provide a description of level of general function, while an extensive list of 25 specific symptoms that can impeded function is also included. Each of the functional levels and symptom is assigned a "preference weight," which then comprises a global quality of well-being score. Preference weighting of this sort has been used as the basis for health care cost-utility analysis, placing relative weighted value upon types of symptom combinations. They are the basis of actuarial systems employing indices such as "quality-adjust life-years," or QALYs.

Psychological Testing

We have saved our discussion of psychological testing as the last topic in our brief and selective overview of psychological assessment activities. This is not because we consider it the least important. Indeed, to the contrary, it is the most important of a psychologist's assessment activities, and we do not consider that any psychological assessment consultation is complete unless, or until, some form of psychological testing has been done and some testing data have been recorded.

Standard of Care

In our discussion of psychological testing we want to introduce several assessment concepts and strategies that have become the norm in clinical practice. The first issue we want to highlight can be labeled "standard of care." The psychologist is no longer an island of practice unto him/herself. Under our almost

universal third-party coverage, insurance companies want to know the quality of care for which they are paying. Similarly, with the rising number of malpractice complaints, both patients and their attorneys want to know what are accepted standards of care and whether the patients' care was at such a level.

Quality of care and patient satisfaction are defined by patient perceptions and expectations, perhaps even more than by standards established by the profession. We have already discussed the importance of doctor–patient communication. Two additional very important components of patient satisfaction are accessibility and availability of services. Our clinic has tried to attend to timely availability to patients and has established and monitored efforts toward this goal.

Timeliness

An important assessment strategy is to respond promptly when a request for a psychological consultation is made. Many inpatients are hospitalized for relatively short periods of time and, if psychological consultation findings and/or recommendations are to be included in decision making, the consultation assessment needs to take place within 24 hours. For outpatient consultations, the patient should also be contacted within 24 hours to arrange a mutually convenient appointment time.

Quality Assurance

In our health care setting, as in most others where hospital or other health facility accreditation is involved, we must show what efforts we make to monitor, evaluate, and improve clinical services. One aspect of such monitoring and evaluation function is to have established standards of care and then to assess whether we are delivering services according to those standards. Standard of care is clearly involved when we consider what constitutes acceptable psychological testing in response to various consultation referrals.

For example, our clinical faculty has established a standard of care in assessing questions of intellectual ability level in patients. When seeing a patient for such a referral question, the core procedures include use of the Wechsler Adult Intelligence Scale—Revised, The Wide Range Achievement Test—Revised-2, and a review of the patient's educational history with a review of school transcripts when these can be obtained. In our Quality Assurance monitoring we review records to confirm that these core procedures have indeed been completed,

Testing Protocols for Different Patient Groups

We have no doubt that, in time to come and in the interests of establishing and assuring appropriate standards of care, testing protocols appropriate to various patient groups, and to various referral questions, will be established. Although practicing psychologists have long discussed the uniqueness of each patient, the advantages of flexible testing, and the desirability of tailoring assessment to each individual, such approaches to assessment raise some important questions in our minds. For example, one cannot know what relationship scores from one test have to scores from another test unless the tests have been used consistently and systematically with each other so that through research or extensive experience such relationships can be identified. When different tests are used with each

patient, it is not possible to observe or establish patterns of test response within a given patient group.

It is also important to observe standard testing procedures. As pointed out by Faust, Ziskin, and Hiers (1991), when a clinician alters standard instructions or test procedures, then short of research on these changes, one does not know how this impacts upon test scores, or what scores would have been achieved had the standard instructions been used.

When consistent test protocols are not used with particular patient groups, or referral questions, it becomes more difficult to determine effectiveness or accuracy across different assessment cases. One has essentially performed a unique set of procedures with every individual, and must conduct a separate "experiment" for each case seen (Faust, Ziskin, & Hiers, 1991). Overall, an approach that allows different assessment procedures among a given diagnostic category of patients, is likely to result in significant variability across different clinicians, is quite difficult to evaluate scientifically, and seems open to examiner bias.

As an assessment strategy in our clinic, we have developed testing protocols for different patient groups. One of these consists of patients referred from the Rheumatology Clinic with a presumed diagnosis of fibromyalgia. Most fibromyalgia patients require some form of psychological intervention, and all of the patients are seen for psychological evaluation when they are accepted into our treatment program (Bennett, Campbell, Burckhardt, Clark, O'Reilly, & Wiens, 1991). We wanted a psychological test battery that would reproduce assessment procedures used by other clinician/investigators and that could be replicated by other clinicians in turn. We chose the following procedures to constitute our protocol: Clinical Interview, C-DIS, Psychological/Social History Questionnaire, MMPI, Cornell Medical Index Health Questionnaire, the Shipley Institute of Living Scale, and a Quality of Life Inventory. We have been able to describe the psychological characteristics of this patient population to a number of different professional audiences.

In concert with clinicians in our Occupational Health Clinic and the Department of Neurology, we believe that behavioral and neurophysiologic changes may be the earliest, and sometimes the only, indicators of acute or chronic neurotoxicity. That is, we view behavior as a sensitive indicator of central nervous system impairment. The literature has suggested that neurotoxins interfere with at least four distinct aspects of central nervous system functions: memory, visuomotor performance, affect, verbal concept formation. We have proposed measurement of behavior by neuropsychological tests for the objective assessment of early neurologic deficit and for the detection of preclinical nervous system changes resulting from environmental or occupational exposure to neurotoxic agents. The core psychological test protocol that we selected for our neurobehavioral examination is as follows: WAIS-R, MMPI, Halstead-Reitan Neuropsychology Test Battery (15 subtests), Rey Auditory Verbal Learning Test, Rey-Osterreith Complex Figure Test, Cornell Medical Index Health Questionnaire, and a Structured Interview. The interview has a number of foci: previous health history, medications, use of alcohol/drugs, present symptoms and history, general orientation, physical appearance and gait, speech, affect, personality characteristics, and general competence to handle daily activities. We have been able to describe the modal cognitive, personality, and physical complaints of these patients and to differentiate among them those who show no identifiable emotional or cognitive dysfunction, those

who show primarily emotional/chacterological concerns and no significant evidence of cognitive impairment, those who have a history of episodes of emotional dysfunction and show evidence of cognitive impairment, and those who have evidence of cognitive impairment as the primary finding. We do not believe that we could have made the observations that we have of this patient group without a standardized testing protocol.

Clinical versus Actuarial Judgment

In this final section we can only introduce discussion of a topic that will preoccupy psychologists increasingly in time to come. In talking about assessment, almost all psychologists would suggest that, in arriving at assessment conclusions the clinician has to "integrate" the assessment data with all of the historical data available. This integrative function is often subsumed under the heading of "clinical judgment." Faust, Ziskin, and Hiers (1991), have challenged us with the following statement:

> Although the need to "integrate all of the data" and the ability to do so are often taken for granted, it is extremely doubtful that clinicians can perform such cognitive operations (p. 279).

Paul Meehl (1954) introduced the issue of clinical versus actuarial judgment to a broad range of social scientists in 1954 and his lucid exposition stimulated a great deal of research on this topic. Dawes, Faust, and Meehl (1989) have more recently reviewed much of this research and have again concluded that research comparing these two approaches shows the actuarial method to be superior. Clinical diagnosticians must be aware of these research findings and face a significant challenge in planning how to incorporate them into clinical practice.

Faust and Ziskin (1988) also addressed the topic of factors limiting clinical judgment. They note, to begin with, that mental health practitioners are limited by the state of their science, in that psychology lacks a formalized, general theory of human behavior that permits accurate prediction. For example they cite the dozens of personality theories and hundreds of approaches to psychotherapy. More specifically, on the point of limitations in clinical judgment, they suggest that clinicians often underutilize information about frequency of occurrence, or base rates. For example, if a suicide indicator occurs in 80% of true cases and 10% of negative cases, and if suicidal intent is present in one per 1000 patients, the one patient is likely to be identified correctly by such a suicide indicator but about 99 will be misidentified. Faust and Ziskin express concern that clinicians often overvalue supportive evidence and undervalue counter-evidence. Clinicians expect to and typically find evidence of abnormality in individuals they examine, even normal persons. Faust and Ziskin also note that clinicians often practice under conditions that do not promote experiential learning; that is, they often receive little or no outcome information or feedback about their judgments. With reference to psychotherapy outcome, it is usually the satisfied patients who may make follow-up contact with the clinician to express their satisfaction. Patients who were unhappy with the clinicians' judgments may simply absent themselves from further contact.

Dawes et al. (1989) make clear that clinical judgment should not be equated with a clinical setting or a clinical practitioner. A clinician in psychiatry, psychol-

ogy, or medicine may use the clinical or actuarial method. The definition of the clinical method is that the decision-maker combines or processes information in his or her head. In the actuarial or statistical method the human judge is eliminated and conclusions rest solely on empirically established relations between data and the condition or event of interest. Dawes et al. go on to say that:

> . . . the actuarial method should not be equated with automated decision rules alone. For example, computers can automate clinical judgments. The computer can be programmed to yield the description "dependency traits," just as the clinical judge would, whenever a certain response appears on a psychological test. To be truly actuarial, interpretations must be both automatic (that is, prespecified or routinized) and based on empirically established relations (1989, p. 243).

Dawes et al. (1989) add that virtually all types of data are amenable to actuarial interpretation. Qualitative observations (e.g., patient appears withdrawn) can be coded quantitatively and incorporated into a predictive equation. Actuarial output statements can be written for virtually any prediction of human interest.

A well-known example of actuarial prediction is the Goldberg Rule in differentiating neurosis from psychosis on the MMPI. (The following research is discussed with an important caveat to the reader: The research was carried out during the years when clinical diagnoses, as noted earlier, were not as reliably made as they are now. It would be of interest to know whether the same findings would result if the research were implemented with present-day clinical diagnostic procedures. Nonetheless, the Goldberg research is of considerable interest.) Goldberg (1965) showed that the most effective rule for distinguishing psychosis from neurosis was quite simple: Add scores from three scales and then subtract scores from two other scales. A cutting score was selected; if the sum falls below 45, the patients is diagnosed neurotic and if above 45 the patient is diagnosed psychotic. The criterion was the patient's discharge diagnosis. The decision rules were then applied to new cases and also compared with clinical judges. In each of seven different settings the Goldberg Rules performed as well as or better than clinical judges. In another study, judges were given training packets and even the outcome of the Goldberg rule for each MMPI, and were free to use the rule when they wished. Judges generally made modest gains in performance, but none could match the rule's accuracy; every judge would have done better by always following the rule. In an interesting elaboration of this research, Goldberg (1970) constructed mathematical models of the judges' decision making. In principle, if a judge weights variables with perfect consistency, the same data will always lead to the same decision and the model will always reproduce the judge's decision. Goldberg found that the judges were not always consistent, and in cases of disagreement the models were more often correct than the very judges on whom they were based.

Dawes et al. (1989) note that the perfect reliability of the models is likely to explain their superior performance in this and related studies. After reviewing a sample of 100 studies that showed the superiority of actuarial decision making in almost every case, Dawes et al. (1989) concluded that the actuarial advantage is general and likely encompasses even judgment tasks not yet studied. They felt that there is no other body of research in psychology in which the findings are coming out as uniformly as they are in the studies of clinical versus actuarial prediction.

In thinking about factors underlying the superiority of actuarial methods, Dawes et al. (1989) note, first of all, that actuarial procedures, unlike the human

judge, always lead to the same conclusion for a given data set. Second, the mathematical features of actuarial methods ensure that variables contribute to conclusions based on their actual predictive power and relation to the criterion of interest. Individuals often have difficulty in distinguishing valid and invalid variables and may develop false beliefs in association between variables. Clinicians often do not obtain immediate feedback on the validity of their diagnoses. Self-fulfilling prophecies may come into play as when prediction of an outcome leads to decisions that influence or bias that outcome. The clinician may also be exposed to a limited or skewed sample of humanity and, without exposure to truly representative samples, may not be able to determine relationships among variables. One cannot determine whether a relation exists unless one also knows whether the sign occurs more frequently among those with, versus those without, the condition. As Dawes et al. (1989) point out, if 10% of brain-damaged individuals make a particular response on a psychological test and only 5% of normals, but nine of ten clinic patients are not brain-damaged, most patients who show the feature will not be brain-damaged.

Although surpassing clinical methods, actuarial procedures are also fallible and sometimes can achieve only modest results. They need to be periodically reevaluated, and they need to be established for each new setting. Reevaluation is aided by the fact that actuarial methods are explicit and can be subjected to informed criticism and be made freely available to other members of the scientific community who might wish to replicate or extend research. Clinician-researchers must lament with Dawes et al. (1989) that the investigations on clinical versus statistical judgment have had so little impact on everyday decision making, particularly within its field of origin (clinical psychology). Although of demonstrated value, actuarial interpretation of interviews is still rarely used. As relevant research findings accumulate, actuarial interpretation will be relied on much more heavily in the future. When actuarial methods prove more accurate than clinical judgment, the benefits to individuals and society are apparent. Much would be gained, for example, by increased accuracy in the prediction of violent behavior and parole violation, the diagnosis of disorder, and the identification of effective treatment (Dawes et al., 1989). Even lacking any outcome information, it is possible to construct models of judges' decision making that will likely surpass their clinical judgment accuracy.

Summary

The assessment strategies that we have been discussing can be expected to lead to a diagnosis of a patient's condition. A great deal of thought has gone into thinking about criteria for diagnoses and sources of unreliability in diagnostic formulations. Spitzer, Endicott, and Robins (1975) noted five sources of unreliability and then determined that two of these contributed most heavily to diagnostic unreliability. The first source of unreliability they noted was *subject variance*, which occurs when patients actually have different conditions at different times. Spitzer et al. gave the example of the patient who may show alcohol intoxication on admission to a hospital but develop delirium tremens several days later. A second source of unreliability is *occasion variance*, which occurs when patients are in different stages of the same condition at different times. An example of this would be a patient with a bipolar disorder, who is depressed during one