Behavioral Measurement of Severe Depression

James G. Williams, MD; David H. Barlow, PhD; and W. Stewart Agras, MD, Jackson, Miss.

Three different methods of measuring depression were studied in ten severely depressed patients. Serial assessments were made using the Beck Depressive Inventory (self-rating scale), the Hamilton Rating Scale (physician-rated scale), and objective behavioral measures. The hypothesis was that these three methods would show a significant correlation; the study confirms this. The behavioral measures are an inexpensive method of assessing the depth of depression and seem to predict the posttreatment adaptation better than the Beck Depressive Inventory or the Hamilton Rating Scale. Behavioral measures should be useful in future research to assess the results of drug or somatic therapies.

Diagnosis and classification of depression has been of interest to clinicians for many years. This interest is justified, for as research methods have become more sophisticated, drug therapies more effective, and biochemical studies more complicated, it has become clear that the validity of most clinical research is dependent upon the homogeneity of the sample of depressed patients. As Mendels and Cochran have pointed out, treating depression as one broad entity may obscure a specific group of depressed patients, create problems of communication, and interfere with comparisons of results between studies.

Before accurate classification or meaningful research can be undertaken, reliable and valid measurement techniques must be developed. Meaningful drug research begins with the selection of homogeneous groups of depressed patients and necessitates the use of accurate, objective measures to assess the results of therapy. The severity or depth of depression has previously been measured by self-rating scales such as the Beck Depressive Inventory or by a physician-rated scale, the Hamilton Rating Scale. Although both these measures are useful, the addition of behavioral measures of depressed behavior should add a new dimension in attempts to quantify subjective feelings of depression.

Few attempts have been made to quantify depression by completely objective behavioral measures. Bunney and Hamburg, in 1963, developed a method for observing and recording behavioral data on a psychiatric ward. These investigators used the nursing personnel on a research unit to observe and record depressed behavior. Their observations rely heavily on verbal behavior and require the nursing staff to make global estimates of mood and feeling at the end of an eight-hour shift. They noted that: "Patients who spoke little and avoided interaction with the staff were difficult to rate."

It was our desire to develop a more objective system of observing and recording depressive behavior that could be administered by a nurse's aide on a general hospital unit.

It was felt that if behavioral measures could be developed, they would provide a precise, yet economical, longitudinal record of the patient's depressed behavior and the response of the depression to drug, somatic, and psychotherapy intervention. This method could supplement assessment by self-report and physician interview, and yet circumvent one of the major disadvantages of the Hamilton Rating Scale in that a physician would not be required to spend time conducting and rating an interview. On the other hand, the pitfalls of self-assessment inherent in the Beck Depressive Inventory would be avoided. To validate the behavioral measures concomitant, sequential evaluations were obtained using the Beck Depressive Inventory and the Hamilton Rating Scale. Our hypothesis was that the behavioral measures and the rating scales would show a positive correlation.

Methods

Selection and Characteristics of the Sample.—This study was conducted on a university hospital psychiatric unit which emphasizes teaching, good clinical care, and research. Diagnoses were made in a morning staff conference by a board-certified psychiatrist, a psychologist, and a second-year psychiatric resident. Data used to establish the diagnoses included information from interviews with the patient and relatives, reports from referring physicians and summaries of previous hospitalizations. These data were organized and presented by a third-year medical student under the supervision of the ward resident. Following the presentation, a social worker's summary was considered and a five- to ten-minute audiovisual tape of the patient was studied. The first ten patients meeting our criteria for the diagnosis of psychotic depression were used as our sample. The following characteristics were used to make the diagnosis:

1. Feelings of hopelessness and worthlessness.
2. Suicidal thoughts or attempts.
3. Loss of interest in life.
4. Self-reproach or guilt.
5. Physiological signs of depression including: (a) anorexia; (b) weight loss; (c) sleep pattern disturbance; (d) impotence; (e) constipation; (f) anergia.
6. Objective signs of deep depression: (a) depressed face; (b) slow monotonous speech; (c) psychomotor retardation.

Measures.—Patients' self-ratings were obtained using the Beck Depressive Inventory. This inventory was developed by Beck et al. in 1961 to assess severity of depression; it was not designed to distinguish between diagnostic categories. The inventory consists of 21 items covering various aspects of depression; each item is subdivided according to severity so that the patient may score 0 to 3 on any item.

The Hamilton Rating Scale was used to record the depth of depression during periodic psychiatric interviews. This scale was designed to assess the depth of depression in patients with a known diagnosis of depressive illness, and should not be used for diagnostic purposes. According to Hamilton: "The scale quantifies the results of an interview and is of practical value in assessing the results of treatment." The scores represent a record of the rater's judgment. Hamilton subdivided ten major categories of depression into 17 items. These items are then individually scored on a...
or 5-point scale by the rater during an interview. The total score ranges from 0 to 100, representing the sum of two raters’ scores, or a double score for one rater.

The development of behavioral measures of depression involved designating overt and observable behavior. Selection of depressed behaviors for observation was dictated by ease of observation, frequency, and constancy. Thus, subtle behaviors, low-frequency behaviors, verbal content, and estimations of affect were avoided. Instead, we sought consistent behavioral cues to depressed affect that were either present or absent without question at any point in time. In this way decisions by the observers were minimized and record keeping simplified, insuring a high degree of reliability. Preliminary observations on our unit suggested that in addition to subtle behaviors, depressed patients talked little, avoided social interaction by withdrawing to their rooms, smiled infrequently, and exhibited diminished motor activity. These measures were defined as follows:

1. **Talking.**—Verbal behavior directed toward another person.
2. **Smiling.**—Facial movement in which the corners of the mouth turn up. The teeth may or may not show.
3. **Motor Activity.**—The following are activities:
   A. Patient in room with visitor or another patient.
   B. Patient at card table talking, reading, or sewing.
   C. Patient in TV lounge talking, reading, or sewing.
   D. Patient in room dressing or straightening up room.
   E. Patient taking shower.
   F. Patient in physical therapy, occupational therapy, or group therapy (or preparing to go).
   G. Patient sitting and watching TV.
   H. Patient at card table or in TV lounge with other patient or patients (even though patient is not actually talking at that particular minute observed).
   I. Patient drinking coffee on eighth floor.
   J. Patient alone in TV room, with set off, reading, sewing, or knitting.

The following are not activities:

A. Patient in room alone—sitting, lying down, or looking out the window.
B. Patient sitting alone in TV room, with set off.
4. **Time Out of Room.**—The patient is not inside the room at the time he is checked.

The behavioral measures are intended to be another means of quantifying the depth of depression; they are not intended to distinguish between diagnostic categories.

**Procedure.**—The Beck Depressive Inventory was administered by psychiatric nurses who had been instructed in its use by one of the investigators (J.G.W.). The inventory was given according to instructions, in the morning and evening every three days throughout the patient’s hospital stay. The nurse read aloud the statements in each category as the patient followed with his copy. When the patient gave his response, the nurse circled the number adjacent to the statement he selected. The nurse had no knowledge of the results we were obtaining with the behavioral ratings or the Hamilton Rating Scale.

Two physicians (second- and third-year psychiatric residents) were familiarized with the Hamilton Rating Scale. Every three days during the patient’s hospitalization, the Hamilton Rating Scale was completed. At first the scores of the two raters were added, but after several trials scores were consistently similar. Then only one resident completed the scale and his score was doubled. The interviewers were blind to the results of the Beck Depressive Inventory and the behavioral ratings.

At the beginning of the study, two training sessions of one hour each were held with two of the aides and a behavioral technician.

The behaviors were defined and rating sheets provided with the definitions. Our aides, who were high school graduates or college students, had no difficulty grasping the observations desired or
were made, averaging one per half hour. These observations were made surreptitiously by the aide assigned to the depressed patient for the day. Within each half hour, however, the times were randomized based on a table of random nurses. For example, a patient might be observed at 10:12 for his 10 to 10:30 observation, and then at 10:40 for his 10:30 to 11 observation. Each morning the aides chose new times from the table of random numbers and noted them on that day’s data sheet. Approximately 30 observations during the hospital course of each patient were recorded by a trained, independent observer in addition to the assigned aide to check on interrater reliability. Interrater reliability obtained in this way was 96%. Separate data sheets were provided each day so that aides were not influenced by the previous day’s recording. Except for electric convulsive therapy, the aides had no knowledge of the patient’s treatment, his score on the Beck Inventory, or his score on the Hamilton Rating Scale.

At weekly intervals raw data sheets were reviewed by a behavioral technician or one of the investigators to avert any errors in recording. Ongoing graphs of the data were plotted by the behavioral technician on each patient currently under investigation. One graph consisted of scores on the Hamilton Rating Scale, Beck Depressive Inventory, and behavioral measures; another graph consisted of the four behavioral measures alone.

**Results**

An analysis of the four behavioral measures was performed with Kendall’s coefficient of concordance. The resulting correlation was highly significant ($W = .70, P < .01$). On the basis of these results the behavioral measures were summed in daily totals and treated as single data on all further analyses.

These behavioral data were then blocked in three-day blocks surrounding the presentation of the Hamilton Rating Scale and the morning administration of the Beck Depressive Inventory. The three days were the day prior to administration of the Hamilton, the day the Hamilton Rating Scale was administered, and the day following administration of the Hamilton Rating Scale. Using the resulting blocked data, Pearson product-moment correlations were computed among the three measures for each patient. For each correlation of two scales, weighted Fisher’s $z$-transformations were made. The mean of the transformed data across patients was then obtained and the resulting value retransformed; this resulted in an overall Pearson product-moment correlation for the two measures. Resulting mean correlations were all significant beyond the .05 level. Overall correlations were:

- Beck and Hamilton: $r = .82$
- Hamilton and behavioral: $r = .71$
- Beck and behavioral: $r = .67$

Certain differences among the behavioral measures, the Hamilton Rating Scale, and the Beck Depressive Inventory emerged. The behavioral measures were sensitive to some environmental influences and physical disabilities of the patients. For example, one of our patients was fearful of Negro men and when a Negro man was admitted, her behavioral score dropped rapidly because she withdrew to her room. However, during the same period her Beck score tended to stabilize as did her Hamilton Rating Scale. Both the Hamilton Rating Scale and the Beck Depressive Inventory scores remained in the nondepressed category while her behavioral measures dropped drastically. When the Negro man was discharged from the unit, her behavioral score rose and tended to parallel the Hamilton Rating Scale score and the Beck Depressive Inventory score. In this patient, self-report helped clarify our behavioral data. Another of our patients was virtually blind, and her overall behavioral score was suppressed somewhat. That is, she tended to have a lower overall baseline behavioral score and her peak behavioral score was lower than the peak score in other patients.

Significantly, the behavioral measures tended to predict the posthospital course more accurately than either of the other two scales. To date, five patients have been followed for one year after discharge. Of these five patients, three have had relapses of their depression. In the hospital their longitudinal behavioral score tended to improve initially.
Platuea medio in their hospital course, and then worsen slightly prior to discharge (Fig 1). The Hamilton Rating Scale and the Beck Depressive Inventory on these patients tended to improve rapidly and top out early, suggesting that they were completely over their depression within the first week or two after hospitalization (this again is evident in Fig 1).

Two patients showed a clear upward trend in their behavioral score prior to discharge; both are well at the end of one year. The Hamilton Rating Scale score and the Beck Depressive Inventory score on both of these patients tended to follow the same pattern as it did in other patients; they improved quite rapidly so that by the third to sixth days of hospitalization both the Hamilton Rating Scale scores and the Beck Depressive Inventory Scores were in the nondepressed category. Figure 2 illustrates these findings.

In another case, the behavioral measures tended to lag behind the Hamilton Rating Scale and the Beck Depressive Inventory, improving much more slowly. In fact, by the sixth hospital day both the Beck Depressive Inventory and the Hamilton Rating Scale reflected that this man was no longer depressed at all. On the third day his Hamilton Rating Scale score was 24 and on the sixth day his Hamilton Rating Scale score was 14. His Beck score on the third day was 2 and he made a score of 0 thereafter on the Beck Depressive Inventory. His behavioral measures began to improve steadily on the sixth day and continued to improve until the ninth day when they plateaued. However, from day 12 until discharge the behavioral measures continued to improve dramatically; his behavioral score more than doubled while his Beck Depressive Inventory score and the Hamilton Rating Scale score remained in the nondepressed level.

This patient’s preadmission course underscores the importance of this finding. Prior to his admission to university hospital his depression was treated at a local hospital by a psychiatrist. A few days after his release from that hospital, the patient cut his throat and antecubital veins in a serious suicide attempt and was admitted to university hospital. Although the circumstances leading to his early discharge are unclear, it seems that in this case the behavioral measures provide a more conservative indicator of progress than the other two measures, allowing longer and more adequate treatment.

Comments

This preliminary study confirms our hypothesis that there is a statistically significant correlation between the Hamilton Rating Scale, the Beck Depressive Inventory, and the behavioral measures in a group of ten severely depressed patients; it indicates that a behavioral assessment of depression is one means of verifying self-report and clinical impression. Since the behaviors are precisely defined and easily observed, no complicated decisions are required by the nursing assistants; thus, little training is required before interrater reliability is established. Consequently, behavioral recording can be accomplished by a nursing assistant with a high school education without interfering with patient care. In fact, as Bunney and Hamburg have noted, careful observation and recording may increase the quality of nursing care.

Although the Hamilton Rating Scale is an accurate instrument for quantifying the degree of depression during a psychiatric interview, its value is dependent on the skill of the interviewer. Consequently, the Hamilton Rating Scale is usually scored by a physician after a 30-minute to one-hour interview. This is not only a more expensive method, but also provides a measure of depression in a structured situation at a specific time of day. Moreover, administration of the Hamilton Rating Scale requires the intervention of an outsider and may be disruptive toward routine.

The use of behavioral methods avoids the pitfalls of self-assessment scales of which the Beck Depressive Inventory is the best known. Aside from the notorious unreliability of self-assessment, the depressive scales are of little use in a seriously ill patient who is unable to deal with them. Although our patients did not refuse to complete the Beck Depressive Inventory, they often protested having to complete it early in their illness or completed it in an apathetic manner, constantly protesting that they could not give the correct response. As patients improved, they completed it rapidly and with seeming unconcern.

Finally, behavioral measures taken over time provide an accurate longitudinal record of behavioral changes during depression. The value of longitudinal observation and rating have been pointed out by Bunney and Hamburg.

Studies can be used to assess the course of a mental illness on a day-to-day basis, to identify periods of crisis or change, and to correlate quantitative estimates with biochemical measurement. In our series, the behavioral measures provided a longitudinal record of the patient’s daily progress and response to treatment. It would be impractical, if not impossible, to administer the Hamilton Rating Scale four to five times a day, and the Beck Depressive Inventory may be unreliable if administered after too short a time because the correlation between the two sets of scores could be spuriously inflated due to memory of previous administration.

Although the behavioral measures may lose some validity due to environmental disruption or possible physical disability, there are indications that behavioral assessment may be more accurate in some cases and predict posttreatment adaptation more successfully than the Hamilton Rating Scale or Beck Depressive Inventory. Confirmation of these findings must await further experimental analysis with a larger sample. Additional research should focus on establishing norms for depressed behavior to facilitate comparisons. In the meantime, this assessment procedure should be a useful tool for research into a little known area of psychopathology.

References