A Brief Rating Scale for Antidepressant Drug Trials

By Burton W. Rockliff

Although a rather wide choice of rating scales is available to psychiatrists and clinical investigators for quantifying symptoms of psychopathology, none of these are completely satisfactory for use in studies of antidepressant drugs. Many of these instruments, such as Inpatient Multidimensional Psychiatric Scale (IMPS) or Brief Psychiatric Rating Scale (BPRS), were not developed primarily as measures of depression. Others, such as the Hamilton Rating Scale, contain many items concerned with anxiety and somatic complaints, thus giving excessive weight to changes produced by the tranquilizing action of the psychotropic agents being tested.

In practice, these rating scales seem to have no advantage over simple global assessments of response, with respect to the primary goal of detecting differences between treatments. In a thorough review of the published reports on double-blind evaluations of imipramine, Klerman and Cole note: "Statistically significant differences between imipramine and control treatment were found most frequently on global ratings of improvement." This coincides with our own conclusions after reviewing the pertinent literature.

Furthermore, some of these scales have so many items to score (some not even remotely related to depression) and are so time-consuming, that frequent ratings during studies are discouraged, and longitudinal data are seldom obtained in drug trials, a deficiency which has been pointed out by others.

Construction of Scale

The simplicity of the Depression Rating Scale devised by Lehmann seemed to offer advantages in this direction, but the scale did not include some important features of depression among its seven items. Using the original scale as a basis, items concerning impairment of work and social interests, suicidal tendencies and somatic complaints were added. Two of the original items, sleep and loss of weight, were modified, and the resultant 10-item scale shown in Table 1 was felt to be fairly inclusive while retaining the brevity and simplicity of the basic instrument.

The ratings of each item are scored on a four-point scale of 0–3, and the item scores are added for the total score, which may range from 0–30. The scale was not intended to be a diagnostic tool, but rather a means of providing a useful measure of changes in the severity of depression in appropriately diagnosed patients under observation for treatment effects. Accordingly, an attempt was made to assess the reliability and validity of the scale by direct
<table>
<thead>
<tr>
<th>Item</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Mood</td>
<td>Severely depressed</td>
<td>Moderately depressed</td>
<td>Mildly depressed</td>
<td>Not depressed</td>
<td></td>
</tr>
<tr>
<td>(2) Facies; general appearance</td>
<td>Severely depressed</td>
<td>Moderately depressed</td>
<td>Mildly depressed</td>
<td>Not depressed</td>
<td></td>
</tr>
<tr>
<td>(3) Psychomotor retardation (observed)</td>
<td>Intense</td>
<td>Moderate</td>
<td>Slight</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>(4) Impairment of work and social interests</td>
<td>Completely inactive</td>
<td>Moderately impaired</td>
<td>Slightly impaired</td>
<td>No impairment</td>
<td></td>
</tr>
<tr>
<td>(5) Agitation</td>
<td>Intense</td>
<td>Moderate</td>
<td>Slight</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>(6) Expressed feelings of guilt, worthlessness, hopelessness, nihilistic delusions</td>
<td>Intense</td>
<td>Moderate</td>
<td>Mild</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>(7) Suicidal tendencies</td>
<td>Overt threats or attempts</td>
<td>Wishes to die</td>
<td>Mild references</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>(8) Insomnia (without hypnotic drugs)</td>
<td>Severe, constant less than 3 hours sleep</td>
<td>Moderate, usually less than 5 hours sleep</td>
<td>Mild sleep disturbance</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>(9) Somatic complaints</td>
<td>Many, unremitting, disabling</td>
<td>Moderate in number, frequency and severity</td>
<td>Few, not disabling</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>(10) Loss of appetite and weight</td>
<td>Completely anorexic; rapidly losing weight</td>
<td>Poor appetite; slowly losing weight</td>
<td>Eats without pleasure; weight stable or increasing</td>
<td>Appetite good; weight stable or increasing</td>
<td>Total</td>
</tr>
</tbody>
</table>
comparison with global assessments in groups of such patients before and during course of treatment.

**Sources of Data**

In a series of evaluations of antidepressant drugs in hospitalized patients, global assessments of the severity of illness in each patient were to be uniformly rated as: 3 = severely depressed, 2 = moderately depressed, 1 = mildly depressed and 0 = not depressed. The investigators were requested to complete the Depression Rating Scale at the same examinations following the global assessments. Two of these studies were straightforward evaluations of a liquid form of imipramine pamoate, without control groups. The third investigation was a double-blind comparison of imipramine hydrochloride and a related compound, ketipramine fumarate. The fourth clinical trial was a double-blind comparison of imipramine pamoate and placebo in a hospitalized group of depressed alcoholics. In the first three studies, global assessments of the severity of depression were made initially (pretreatment) and at weekly intervals during treatment. Only the initial assessments followed these same criteria in the fourth investigation; global ratings during treatment were based on improvement rather than severity, and these data are analyzed separately.

The global assessments of severity and the simultaneously obtained rating scale scores from these four studies have been pooled for the analysis. Both initial and interval examinations contributed global ratings of 1, 2, and 3 to the pooled data. Only patients considered recovered on interval assessments received a global rating of “not depressed” (0).

The patients in the fourth study were examined at the end of 1, 2, and 3 weeks of treatment, and their response was rated very much improved, much improved, slightly improved, no change or worse, as compared to the initial pretreatment global assessment of severity. The rating scale was scored at the same examination.

Finally, the Depression Rating Scale was used in a fifth investigation which compared the effects of desipramine hydrochloride and placebo in newly hospitalized, nonschizophrenic depressives. All ratings were done by two psychiatrists, each of whom performed all the clinical assessments on any single patient. The maximum period of observation was 4 weeks, and ratings were obtained on the day treatment began and at 3, 5, 7, 10, 14, 21, and 28 days. At the initial examination, individual symptoms of depression were noted and scored for severity. At subsequent intervals, a global assessment of change from the initial state was made and rated as complete remission (3+), marked improvement (+2), moderate improvement (+1), unchanged (0) or worse (−1). The rating scale was scored at the initial examination and at each interval during treatment.

**Method of Analysis**

Of the requirements for a depression rating scale listed by Wechsler et al, replicability was not tested directly for a number of practical and theoretical reasons which will be dis-
cussed. The main effort has been directed at validation of the scale through several types of comparisons with global ratings of either severity or improvement.

If the scale is valid, it should differentiate between depressed patients judged to have varying levels of severity before and/or during treatment. This was tested by analysis of the pooled data for the four severity categories based on global assessments. The mean total scores of the Depression Rating Scale for each severity class were compared, and the difference between means of adjacent groups was tested for significance by t test.

Differentiation of varying degrees of improvement at the end of treatment has been considered a requirement for validation of such a scale. This has been carried a step farther in this analysis by comparing the rating scale scores with global estimates of improvement at three intervals during treatment in one study, and at seven treatment intervals in a second study. Correlations have been shown by simple plotting of group means.

Finally, and of greatest importance in our view, the sensitivity of the rating scale has been tested by examining how well it differentiates between responses of two treatment groups when compared with global ratings of response. In the two placebo-controlled studies, intergroup differences at each interval were tested for significance by analysis of variance (ANOVA) or analysis of covariance (ANOCOVA) for the rating scale scores, and by chi-square for the global ratings.

Because the last study provided individual item scores and total scores in a large group of carefully diagnosed patients, some additional analyses were carried out by diagnostic groups and by mean item scores.

The results of both placebo-controlled studies are being published in reports that provide the details of methods, dosages and patient demography.

**RESULTS**

The relation of rating scale scores to the global ratings of severity of depression is shown in Table 2. The data consisted of 529 separate ratings on a total of 165 patients. Differences between means of adjacent severity classes were reasonably similar (8.29; 7.16; 6.09) and t tests indicated that all three differences were highly significant (p<0.01).

Examination of the raw total scores of patients judged to be “not depressed” revealed that 81 of the 88 ratings were scored 2 or less. No score was higher than 4 for this class. Thus, in 92% of cases in which patients were felt to have recovered, the scores fell below 3. Furthermore, only seven out of the 121 scores for the “mild” assessments were less than 3 (actually, all seven scores were 2), so that 95% of such cases had scores of 3 or higher. An estimate of the practical borderline between “not depressed” and “mildly depressed,” therefore, can be considered 2 or less and 3 or more, respectively. The other two borders between severity classes would be more difficult to define, and have no par-

<table>
<thead>
<tr>
<th>Class</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe</td>
<td>156</td>
<td>164</td>
<td>121</td>
<td>88</td>
</tr>
<tr>
<td>Moderate</td>
<td>22.57</td>
<td>14.28</td>
<td>7.12</td>
<td>1.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Differences</th>
<th>Class A</th>
<th>Class B</th>
<th>Class C</th>
<th>Class D</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>8.29</td>
<td>7.16</td>
<td>6.09</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.—Rating Scale Mean Scores by Severity of Depression from 529 Pooled Ratings of 165 Patients
Fig. 1.—Mean global improvement ratings (upper curves) and mean rating scale scores (lower curves) of two treatment groups (N = 20) of depressed alcoholics during withdrawal period.8

![Graph showing mean global improvement ratings and mean rating scale scores over weeks.]

ticular practical value. Further analyses of these pooled data were not deemed worthwhile.

The improvement ratings in the fourth study were given numerical values as follows: very much improved = 3; much improved = 2; slightly improved = 1; no change = 0; and worse = −1 (there was only one rating in this last category). These values were used only for obtaining means to plot the recovery curve shown in Fig. 1. Differences between global assessments of improvement by intervals or groups were tested by chi-square using the original response groupings.

The rating scale mean scores and the mean global improvement ratings for both treatment groups of 20 patients each are shown at the start and at three weekly intervals in Fig. 1. The similarity of the paired curves is obvious.

After 1 week, the mean scores of both treatment groups showed changes from the baseline mean scores which were significant as tested by ANOVA and t test. The changes and significance levels are shown in Table 3. At 2 and 3 weeks, the changes from baseline were significant in all cases beyond the 0.0001 level.

The differences between treatment groups were not significant at 1 and 2 weeks, using either the rating scale scores or the global ratings as measures. At 3 weeks, the difference between rating scale mean scores for the two groups was 1.88 and this difference was of borderline significance (ANOVA p = 0.06; t test p = 0.066). The difference in distributions of improvement ratings at this interval was also at a borderline level of significance (chi-square p = 0.069). The two rating methods provided comparable sensitivity in detecting modest treatment differences in this small-sample study.

In the fifth study comparing the effects of desipramine and placebo a more elaborate analysis was performed. Because some patients improved early and were discharged from the hospital before the completion of 4 weeks of treatment, data for the later rating intervals were available from smaller patient samples. A total of 74 patients completed at least 1 week of treatment (minimum of four ratings), 34 on desipramine and 40 on placebo. The initial (pretreatment) rating scale mean scores were 22.47 for the drug-treated group and 22.75 for the control group.

The rating scale mean scores at all intervals for both treatment groups are indicated in Fig. 2. The mean scores decreased more rapidly in the drug-treated group, the differences becoming significant, as tested by ANCOVA, on day 7 (p < 0.02) day 10 (p < 0.014) and day 14 (p < 0.002). By the 21st day, five recovered patients had been lost from the desipramine group while one placebo-treated patient with poor response had been removed from the study and given ECT. The intergroup difference at this interval was of borderline significance (p < 0.07), therefore, and further attrition reduced the difference to a nonsignificant level by day 28.

Global assessments of change were treated numerically (+3, +2, +1, 0, −1)

Table 3.—Rating Scale Mean Scores Before and After 1 Week of Treatment in 40 Depressed Alcoholic Patients

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>1 Week</th>
<th>Difference</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N  Mean</td>
<td>N Mean</td>
<td></td>
<td>ANOVA/t test</td>
</tr>
<tr>
<td>Imipramine</td>
<td>20 16.05</td>
<td>20 10.50</td>
<td>5.55</td>
<td>p &lt; 0.0001</td>
</tr>
<tr>
<td>Placebo</td>
<td>20 15.65</td>
<td>20 10.95</td>
<td>4.70</td>
<td>p &lt; 0.001</td>
</tr>
</tbody>
</table>

Fig. 2.—Improvement curves by mean rating scale scores (top) and mean global changes (bottom) in a double-blind antidepressant drug trial.8 Numbers in parentheses show the sample sizes initially and at each interval. Significant intergroup differences are indicated by vertical lines.
for the purpose of plotting the improvement curve in Fig. 2. The similarity of the curves of the two treatment groups to the corresponding curves derived from the Depression Rating Scale scores is apparent. Intergroup differences in distribution of global improvement ratings were tested by chi-square. Because there were only three instances in which a rating of "worse" (-1) occurred, this classification was combined with the "no change" rating for this analysis.

The only interval at which the global ratings of the two treatment groups were significantly different was the 10th day (p<.05). Thus, significant between-treatment differences were detected by rating scale scores at three intervals, but global ratings demonstrated a statistically significant difference at only one of these intervals.

When only the endogenous depressives in each treatment group were examined in the same manner, the results demonstrated the interesting fact that all of the superiority in the response of the drug-treated patients was accounted for by the better performance of this diagnostic subgroup. The improvement curves by rating scale scores and by global assessments of change are shown in Fig. 3.

In the 49 patients with endogenous depression, differential changes in rating scale scores of the two treatment groups reached a statistically significant level (p<0.01) as tested by ANCOVA on day 5, and the differences remained highly significant for the next four intervals at the following levels:

day 7 p<0.001; day 10 p<0.0002; day 14 p<0.0004; day 21 p<0.008. Intergroup differences in global improvement ratings were significant at four intervals (chi-square): day 5 p<0.05; day 10 p<0.005; day 14 p<0.005; day 21 p<0.01.

The improvement curves for the remaining 25 patients with neurotic depression are shown in Fig. 4. While the patients on placebo appear to have responded somewhat more completely than the drug-treated patients, none of the differences at any interval were significant for either method of rating. Comparison of the changing scores of these neurotic depressives with those of patients with endogenous depression (Fig. 3) shows that the slower improvement of the control group in the latter patients was more important in producing the significant drug-placebo difference than the somewhat better improvement rate of the endogenous patients on drug over their neurotic counterparts.

Because of the striking difference in results between these two diagnostic groups, an examination of the initial ratings by item scores was done. The mean total score for the endogenous group (23.7) was somewhat higher than that of the neurotic patients (20.5), and when the mean item scores for the two groups are displayed as profiles as in Fig. 5, this difference in severity is found to occur in almost every item. The differences as tested by chi-square were significant at 0.01 and 0.02 levels for items two (appearance), three (psychomotor retardation), four (impaired work and interests) and five (agi-
Approximately 60–70% of the endogenous depressives were rated severe on these items compared to 20–35% of the neurotics.

To investigate the possibility that the differences in response may have been related to the initial severity of depression rather than the diagnostic groupings, the cases were divided into "high initial scorers" (total score 22 or more, N = 41) and "low initial scorers" (total score 21 or less, N = 33). The "high" group had a mean score of 25.7, the "low" group 18.8. Ten of the 25 neurotic depressives (five on drug, five on placebo) were included in the "high" group. In both groups the drug-treated patients responded better on the average than those on placebo. The differences in response between treatments in the more severely depressed group were significant (p < 0.01) on days 7, 10, 14, and 21. In the less severely depressed group, the difference reached borderline significance (p = 0.06) on the 14th day at which time only 13 patients were in the drug-treated subgroup.

The low initial scores in both diagnostic groups for items nine (somatic complaints) and 10 (appetite/weight loss) might suggest that these factors should be "weighted," or possibly eliminated from the scale as noncontributory. The best pragmatic criterion for the usefulness of any item in this scale, however, is its contribution to the differentiat1 between treatments. In order to examine this, between-treatment comparisons of all 10 item scores by total treatment groups and by the two diagnostic subgroups were done, with differences in the distributions of severity scores tested by chi-square. As could be expected from the total score data, there were no significant differences found in the neurotic subgroups. The significant differences in the total groups and the endogenous subgroups are indicated in Tabl4.

While each of these p values need not be accepted literally because of the large number of significance tests performed (some random differences at the 0.05 or 0.01 levels might be expected) the figures do indicate that after 5–10 days differential response to the two treatments became evident in all but one symptom area. Only item seven (suicidal tendencies) failed to show a difference at significant levels between treatment groups at any interval.

**DISCUSSION**

In some respects, the procedures used by Wechsler et al.10 in the validation of their scale have been followed here. The main departure is the lack of any attempt to test inter-rater reliability with this Depression Rating Scale. The relatively poor correlations obtained by those authors provided a strong argument for ignoring this procedure. The interval between the two independent ratings on the same patient had a marked effect on the correlations of the scores, ranging from r = 0.52 when both ratings were done on the same day, to r = 0.78 when the ratings were 1 week apart. The problem is that the response of the patient to the second interview is influenced by the first interview when the two ratings are done within a short time interval. If a longer time elapses between the ratings, the patient's condition may have changed, invalidating any comparison of inter-rater agreement. If both raters conduct the interview together, then the correlation is greatly improved, but the high degree of agreement is likely to be a result of mutual cueing rather than independent judgments based on the response of the patient alone.

While there are techniques which minimize these difficulties, such as using a standard filmed interview to test several raters, the question of inter-rater reliability was felt to be of marginal importance. Judgments about the presence or severity of symptoms will differ among raters due to a host of factors including the length of the interview, the amount of probing, previous observations of the patient and so forth. Unless the rating sessions are rigidly structured and controlled, one can hardly expect a simple rating scale to eliminate these differences among raters. If only crude judgments are required, of course, replicability is enhanced. Because this scale was designed to be used in antidepressant drug trials, however, sensitivity to small changes in the status of the patient was a primary goal, and our main efforts were directed to evaluating the performance of the instrument with respect to this property.

The analysis of the 529 ratings in 165 patients is somewhat crude in that these are not independent ratings. If only initial ratings of these hospitalized patients were used, however, few ratings in the "mild" class, and none in the "not depressed" class would have been available. The data have been used primarily to obtain rough ranges of rating scale scores for patients in each of the three severity categories or judged to have no depression.

The pooled rating scores and global assessments of 165 patients were ob-
tained by three psychiatrists at three different hospitals at widely separated locations. The patient populations probably differed significantly, with one group composed entirely of alcoholic patients in withdrawal phase. The investigators were given no specific instructions about the use of the rating scale, the assumption being made that the items and scoring were self-explanatory. Under these circumstances, the correlations between severity classifications and rating scale total scores were quite satisfactory. The close cluster of low score of those patients considered "not depressed" was particularly gratifying because this is an end-point on which agreement among raters is of considerable practical importance.

The value of the rating scale was demonstrated most effectively in the placebo-controlled evaluation of desipramine. Because the neurotic depressives in that study responded about the same to both treatments, these patients actually diluted the contrast between treatment effects when only the results in the total treatment groups were analyzed. Despite this dilution factor, which is probably present in most studies of heterogeneous samples of depressed patients, significant differences between treatments were detected by rating scale scores at three intervals, as compared to the single significant difference provided by global ratings. Whenever both rating methods demonstrated significant differences at the same interval, the significance level of the rating scale difference was consistently higher than that shown by global ratings. The rating scale scores thus provided a more consistent and sensitive measure of changes in severity of depression than the global improvement ratings.

These results are similar to the experience with this rating scale recently reported by Karkalas and Lal in a placebo-controlled evaluation of imipramine pamoate. In a group of 39 hospitalized depressives, rating scale scores after 4 weeks of treatment showed a difference between treatments at a significant level, although this was not demonstrated by overall clinical response.

With regard to the individual symptom factors, the high scores for both "psychomotor retardation" and "agitation" initially in the patients with endogenous depression are of interest. These symptoms responded equally well to antidepressant therapy, and to a significantly better degree than in the control group. This casts doubt on the validity of separating depressed patients into "retarded" and "agitated" categories for the purpose of choosing a "proper" drug regimen. This concept may be too simplistic and based more on the therapist's preconceptions and distaste for diagnosis than on empirical results.

Perhaps the lack of a clear distinction between agitation and anxiety lies at the root of the matter. Items related to a anxiety, tension, and fears have been omitted purposely from the Depression Rating Scale. While these symptoms may often accompany depression, particularly the neurotic variety, they are not part of the "core of depressive symptomatology" and are not targets of the purely antidepressant actions of drugs. When drugs with combined anti-anxiety and antidepressant activity are evaluated, these symptoms can be rated separately, or one may use the Hamilton Depression Rating Scale which includes a number of items relevant to anxiety. Where a large proportion of the items measure anxiety, as in the Hamilton scale, however, the analysis of total scores alone in drug trials may be deceptive, because anxious agents without antidepressant activity might produce significant score changes, particularly in neurotic depressives. When such scales are used, the scores of depression-relevant and anxiety-relevant items should be analyzed separately.

Despite the fact that the symptom factor suicidal tendencies (item seven) did not show significant between-treatment differences in the placebo-controlled evaluation of desipramine, the rather high initial scores for this item in both diagnostic groups demonstrates the clinical importance of the symptom, particularly in patients considered so ill as to require hospitalization. This symptom may become less observable in a hospital setting, regardless of treatment. There was a definite trend to more prompt disappearance of suicidal thoughts in the drug-treated endogenous depressives, and the deletion of this clinically meaningful item from the scale on the basis of the present data is not felt to be justified.

While all the data used in this analysis were obtained from examinations of hospitalized depressed patients, the Depression Rating Scale has also been used for evaluations of outpatients with depression. The mean total score of a group of 23 depressed female office patients seen primarily for gynecologic complaints was 18.8 prior to treatment with an antidepressant. The rating scale scores of these patients correlated fairly well with the total scores from a self-administered questionnaire concerned with depression (SRQ-D). Fink et al. used the Depression Rating Scale to evaluate severity of depression in 19 ambulatory clinic patients who had failed to respond to imipramine. Marked or moderate global improvement in 10 of these patients on cyclohexone corresponded with decreases in scores of all 10 symptom items, with depressed appearance showing the greatest degree of improvement and insomnia the least. The profile of symptom factors before treatment in this group showed a low score for psychomotor retardation and a rather high insomnia score, but otherwise the ranking of items by severity was similar to that for the group of endogenous depressives shown in Fig. 5.

**Summary**

A 10-item Depression Rating Scale was used to rate the severity of depression of 165 patients before and during treatment in four separate antidepressant drug trials. Global assessments of the severity of depression were made simultaneously. Comparison of 529 paired ratings showed a good relation between total scores of the rating scale and the global severity classifications, with statistically highly significant differences between mean scores for adjacent severity classes.

In two placebo-controlled studies of antidepressant drugs, scores of the Depression Rating Scale were compared with simultaneously obtained global ratings of response. Improvement curves by the two rating methods were similar in both studies. Differences between treatments in one study were demonstrated equally well by rating scale scores and global ratings. In the
Fig. 5.—Mean initial scores of each rating scale item for the endogenous (N = 49) and neurotic (N = 25) depressives combined from the treatment groups shown in Fig. 2. The mean total score for each diagnostic group divided by 10 (thus reduced to the same scale as the item scores) is shown by the dashed line for the neurotic group and the dotted line for the endogenous group.

second study, statistically highly significant between-treatment differences were demonstrated by rating scale scores at three intervals in the total groups and five intervals in the endogenous subgroups, as compared with one interval and four intervals, respectively, by global ratings. Significance levels of differences by rating scale scores were consistently higher than the differences found in global ratings. Because of its brevity, ease of use, and sensitivity, the Depression Rating Scale was found to be a useful instrument for measuring changes in the severity of depression in antidepressant drug trials.

ACKNOWLEDGMENTS

I am indebted to Dr. L. N. Doyle, Oregon State Hospital, Dr. A. T. Butterworth, East Louisiana State Hospital, and Dr. Anthony Lapolla and Dr. Harry Jones, Camarillo State Hospital (Calif.) for their cooperation in testing the rating scale and for providing the data for analysis. Mrs. Bonnie Frazier and Mr. Stuart Blickstein supervised the separate computer analyses of data from the fourth and fifth studies, respectively. Dr. Donald Boyer, Mrs. Sandra Senior and Mr. Percy Tilton provided statistical assistance. The helpful suggestions of Dr. Heinz Lehmann at the time this modification of his rating scale was devised are gratefully acknowledged.

REFERENCES